

# C&S SIG

---

## ***INCORPORAÇÃO DE INCERTEZA NA ELABORAÇÃO DE BASES DE DADOS DE REFERÊNCIA PARA AVALIAÇÃO DA EXACTIDÃO TEMÁTICA DE MAPAS DE OCUPAÇÃO DO SOLO***

---

Pedro Alexandre Reis Sarmento

---

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Ciência e Sistemas de Informação Geográfica

---

Instituto Superior de Estatística e Gestão de Informação  
da Universidade Nova de Lisboa

**INCORPORAÇÃO DE INCERTEZA NA ELABORAÇÃO DE  
BASES DE DADOS DE REFERÊNCIA PARA AVALIAÇÃO DA  
EXACTIDÃO TEMÁTICA DE MAPAS DE OCUPAÇÃO DO  
SOLO**

Dissertação orientada por  
Professor Doutor Mário Caetano  
e  
Professora Doutora Cidália Fonte

Julho de 2009

## **AGRADECIMENTOS**

Aos meus pais por todo o apoio prestado na elaboração desta tese. Sem o vosso apoio incondicional esta tese nunca teria sido realizada.

À Xana, pelo tempo que não passei contigo para elaborar esta tese.

Ao Professor Mário Caetano, pela confiança, amizade, disponibilidade e oportunidade que me tem dado, para desenvolver investigação no Grupo de Detecção Remota do Instituto Geográfico Português.

Ao Hugo Carrão, por toda a amizade, apoio, e ideias que foram essenciais na elaboração desta tese. Esta tese também é tua....

À Professora Cidália Fonte pela sua colaboração e apoio como co-orientadora. O seu apoio foi fundamental e espero que continue a ser no desenvolvimento das linhas de investigação desta tese.

Ao Professor Stephen V. Stehman pelo enorme conhecimento e sugestões que imprimiu em muitas das ideias que estão nesta tese e que deram a confiança necessária de que estamos no caminho certo na investigação que estamos a desenvolver no Grupo de Detecção Remota do Instituto Geográfico Português.

A todos os colaboradores do Grupo de Detecção Remota (Tó, António, Conceição, Vasco) pela vossa disponibilidade e ajuda em todos os momentos. O espírito de trabalho e entre ajuda não podia ser melhor....

# **INCORPORAÇÃO DE INCERTEZA NA ELABORAÇÃO DE BASES DE DADOS DE REFERÊNCIA PARA AVALIAÇÃO DA EXACTIDÃO TEMÁTICA DE MAPAS DE OCUPAÇÃO DO SOLO**

## **RESUMO**

Tradicionalmente, a avaliação da exactidão temática de mapas de ocupação do solo é baseada numa amostra de referência seleccionada aleatoriamente sobre a área em estudo. Nesta abordagem é assumido que a amostra de referência, na qual se pretende representar a verdadeira ocupação do solo sobre a superfície terrestre, não possui erros, facto este que não se verifica. Muitas vezes os erros dos mapas são devidos à incerteza que existe na escolha da classe de ocupação mais adequada para caracterizar uma determinada área, através da foto-interpretação de imagens aéreas de elevada resolução espacial e/ou visitas de campo e/ou devidos ao não co-registo perfeito entre os dados do mapa e os dados de referência. Esta incerteza é devida à fragmentação da paisagem e à existência de mais do que uma classe de ocupação do solo nas áreas amostradas, como também surge da dificuldade na identificação dos limiares de percentagem de coberto utilizados na definição das classes da nomenclatura. De forma a incorporar a incerteza na escolha da ‘verdadeira’ classe de ocupação do solo e a incerteza posicional, os métodos de recolha da informação de referência podem incluir a recolha de mais do que uma classe de ocupação do solo, com um índice de confiança na interpretação, bem como a recolha de um índice de confiança baseado na localização, que reflecte as características da paisagem em cada observação da amostra. Esta informação adicional é utilizada para elaborar duas variáveis nominais (denominadas de CONF e LOC), em que as suas classes reflectem a confiança que se pode ter na classificação do mapa em cada uma das observações de referência. Duas medidas de exactidão temática são posteriormente derivadas, através de pesos atribuídos a cada uma das categorias das variáveis CONF e LOC. A metodologia proposta é testada recorrendo a um mapa de Portugal Continental. Com a metodologia proposta pretende-se complementar a análise tradicional de avaliação da exactidão temática, de modo a permitir uma melhor compreensão da exactidão temática de mapas de ocupação do solo.

# **INCORPORATING REFERENCE CLASSIFICATION UNCERTAINTY IN THE ACCURACY ASSESSMENT OF LAND COVER MAPS**

## **ABSTRACT**

The accuracy assessment of land cover maps is traditionally based on reference sample observations randomly selected over the study area. It is assumed that reference sample observations, representing the “real” land cover at Earth’s surface, are free of errors. However, some of these may be erroneous. These errors are sometimes due to an uncertainty in the identification of the most adequate reference land cover classes by visual interpretation of aerial images and/or field work and/or due to an imperfect registration between the map and reference data. This uncertainty is caused by landscape fragmentation and/or presence of more than one land cover class in sampled areas, and also because difficulties in the identification of the threshold of coverage that is defined for each land cover class of the nomenclature. To accommodate the difficulty of identifying a single ‘true’ or ‘reference’ class and the positional uncertainty, the reference data protocol of an accuracy assessment may include identifying both a primary and alternate reference land-cover label along with a rating of the interpreter’s confidence and also a location confidence rating that reflects the landscape characteristics in the reference classification obtained for each sample location. This additional reference information is used to construct two nominal variables (called CONF and LOC) in which the categories represent the ‘confidence’ in the correctness of the map land-cover classification at a given location. Two accuracy measures that incorporate uncertainty in the reference classification are then derived by assigning partial credit weights to each CONF and LOC classes. The analysis is illustrated using an accuracy assessment of a land-cover map of Portugal. These analyses incorporating uncertainty in the reference classification are intended to supplement traditional analyses to further enhance understanding of the accuracy of land-cover maps.

## **PALAVRAS-CHAVE**

Mapas de ocupação do solo  
Avaliação da exactidão temática  
Bases de dados de referência  
Incerteza temática  
Incerteza posicional

## **KEYWORDS**

Land cover maps  
Accuracy assessment  
Reference database  
Thematic uncertainty  
Positional uncertainty

## ACRÓNIMOS

**CEOS** - *Committee on Earth Observation Satellites*

**CORINE** – *Co-Ordination of Information on the Environment*

**CLC** – *CORINE Land Cover*

**DGRF** – Direcção Geral dos Recursos Florestais

**ESA** – *European Space Agency*

**GDR** – Grupo de Detecção Remota

**GLC** – *Global Land Cover*

**GOFC-GOLD** - *Global Observation for Forest and Land Cover Dynamics*

**ICR** – *Interpretation Confidence Rating*

**IGBP** - *International Geosphere-Biosphere Programme*

**IGP** – Instituto Geográfico Português

**ISEGI** – Instituto Superior de Estatística e Gestão de Informação

**L1** – Classe de ocupação do solo principal

**L2** – Classe de ocupação do solo secundária

**LCR** – *Location Confidence rating*

**MERIS** - *Medium Resolution Imaging Spectrometer Instrument*

**NASA** - *National Aeronautics and Space Administration*

**UNL** – Universidade Nova de Lisboa

**WGCV** - *Working Group on Calibration & Validation*

## ÍNDICE DO TEXTO

AGRADECIMENTOS .....	iii
RESUMO.....	iv
ABSTRACT .....	v
PALAVRAS-CHAVE .....	vi
KEYWORDS.....	vi
ACRÓNIMOS .....	vii
ÍNDICE DO TEXTO .....	viii
ÍNDICE DE TABELAS .....	x
ÍNDICE DE FIGURAS .....	xiv
1. INTRODUÇÃO .....	1
1.1 Avaliação da exactidão temática de mapas de ocupação do solo .....	2
1.2 Limitações da abordagem tradicional de avaliação da exactidão temática de mapas de ocupação do solo.....	4
1.3 Incerteza na elaboração de bases de dados de referência .....	5
1.4 Motivações.....	11
1.5 Objectivos .....	12
1.6 Estrutura da dissertação .....	13
2. METODOLOGIA.....	14
2.1 Base de dados de referência.....	14
2.2 Avaliação da exactidão temática através dos métodos tradicionais .....	16



2.3 Incerteza da referência na avaliação da exactidão temática .....	18
2.4 Cálculo das medidas de exactidão temática $P_{CONF}$ e $P_{LOC}$ .....	19
2.4.1 <i>Elaboração do esquema de pesos</i> .....	21
3. CASO DE ESTUDO.....	25
3.1 Nomenclatura.....	25
3.2 Mapa de ocupação do solo para Portugal Continental.....	27
3.3 Base de dados de referência.....	29
3.4 Resultados.....	31
3.4.1 <i>Análise de sensibilidade</i> .....	41
4. CONCLUSÃO .....	44
REFERÊNCIAS .....	47

## ÍNDICE DE TABELAS

TABELA 1. ESCALA LINGUÍSTICA E RESPECTIVA DESCRIÇÃO UTILIZADA NO TRABALHO DE GOPAL E WOODCOCK (1994) (ADAPTADO DE GOPAL E WOODCOCK 1994).....	7
TABELA 2. INFORMAÇÃO RECOLHIDA EM CADA UMA DAS OBSERVAÇÕES DA AMOSTRA DE REFERÊNCIA (ADAPTADO DE ZHU <i>ET AL.</i> 2000).....	15
TABELA 3. CATEGORIAS CONF AGRUPADAS POR ACERTOS E ERROS ENTRE A CLASSIFICAÇÃO DO MAPA E A CLASSIFICAÇÃO DE REFERÊNCIA. O ACORDO ENTRE AMBOS OS CONJUNTOS DE DADOS É ESPECIFICADO POR ( $\surd$ ). ....	18
TABELA 4. CATEGORIAS LOC AGRUPADAS POR ACERTOS E ERROS ENTRE A CLASSIFICAÇÃO DO MAPA E A CLASSIFICAÇÃO DE REFERÊNCIA. O ACORDO ENTRE AMBOS OS CONJUNTOS DE DADOS É ESPECIFICADO POR ( $\surd$ ). ....	19
TABELA 5. CATEGORIAS CONF E RESPECTIVAS CONTRIBUIÇÕES PARA O CÁLCULO DA EXACTIDÃO DE $P_{CONF}$ . O ACORDO ENTRE AMBOS OS CONJUNTOS DE DADOS É ESPECIFICADO POR ( $\surd$ ). $W_{D,CONF}$ , $W_{D,MAX}$ E $W_{D,RIGHT}$ EXPRESSAM AS CONTRIBUIÇÕES (PARCIAIS OU TOTAIS) PARA O CÁLCULO DA EXACTIDÃO DE $P_{CONF}$ , MAX E RIGHT RESPECTIVAMENTE. ....	22
TABELA 6. CATEGORIAS LOC E RESPECTIVAS CONTRIBUIÇÕES PARA O CÁLCULO DA EXACTIDÃO DE $P_{LOC}$ . O ACORDO ENTRE AMBOS OS CONJUNTOS DE DADOS É ESPECIFICADO POR ( $\surd$ ). $W_{D,LOC}$ , $W_{D,MAX}$ E $W_{D,RIGHT}$ EXPRESSAM AS CONTRIBUIÇÕES (PARCIAIS OU TOTAIS) PARA O CÁLCULO DA EXACTIDÃO DE $P_{LOC}$ , MAX E RIGHT RESPECTIVAMENTE. ....	22
TABELA 7. CATEGORIAS CONF E RESPECTIVAS CONTRIBUIÇÕES PARA O CÁLCULO DE $P_{CONF}$ DEFINIDAS POR UM UTILIZADOR INTERESSADO EM QUE A CONFIANÇA NA CLASSIFICAÇÃO DO MAPA SEJA PELO MENOS ACEITÁVEL. $W_{D,CONF}$ EXPRESSA AS CONTRIBUIÇÕES (PARCIAIS OU TOTAIS) PARA O CÁLCULO DA EXACTIDÃO DE $P_{CONF}$ .....	24

TABELA 8. CATEGORIAS LOC E RESPECTIVAS CONTRIBUIÇÕES PARA O CÁLCULO DE $P_{LOC}$ DEFINIDAS POR UM UTILIZADOR INTERESSADO EM QUE A CONFIANÇA NA CLASSIFICAÇÃO DO MAPA SEJA PELO MENOS ACEITÁVEL. $W_{d,LOC}$ EXPRESSA AS CONTRIBUIÇÕES (PARCIAIS OU TOTAIS) PARA O CÁLCULO DA EXACTIDÃO DE $P_{LOC}$ . .....	24
TABELA 9. NOMENCLATURA DE OCUPAÇÃO DO SOLO LANDEO (FONTE: CARRÃO <i>ET AL.</i> 2009).....	25
TABELA 10. NÚMERO DE OBSERVAÇÕES RECOLHIDAS PARA CADA CLASSE DE OCUPAÇÃO DO SOLO, COMO OCUPAÇÃO DO SOLO PRINCIPAL (L1) E SECUNDÁRIA (L2). ....	31
TABELA 11. ÍNDICES DE CONFIANÇA DE INTERPRETAÇÃO MÉDIOS (ICR) PARA OS DADOS DE REFERÊNCIA, E ACERTOS E ERROS ENTRE MAPA E REFERÊNCIA.....	32
TABELA 12. ÍNDICES DE CONFIANÇA DE LOCALIZAÇÃO MÉDIOS (LCR) PARA OS DADOS DE REFERÊNCIA, E ACERTOS E ERROS ENTRE MAPA E REFERÊNCIA. ....	33
TABELA 13. PROPORÇÃO DE OBSERVAÇÕES EM CADA CATEGORIA CONF POR CLASSE DE OCUPAÇÃO DO SOLO, E AS MEDIDAS DE EXACTIDÃO TEMÁTICA MAX, RIGHT E $P_{CONF}$ . AS MEDIDAS DE EXACTIDÃO APRESENTADAS, SÃO ANÁLOGAS À EXACTIDÃO DO UTILIZADOR, PORQUE O DIVISOR É BASEADO NO NÚMERO DE OBSERVAÇÕES PRESENTES NO MAPA PARA CADA CLASSE DE OCUPAÇÃO DO SOLO. OS ACRÓNIMOS DAS CATEGORIAS CONF SÃO: E3=ERRO MUITO ELEVADO; E2=ERRO ELEVADO; E1=ERRO ACEITÁVEL; A1=MUITO REDUZIDA CONFIANÇA NA CLASSIFICAÇÃO; A2=REDUZIDA CONFIANÇA NA CLASSIFICAÇÃO; A3=ACEITÁVEL CONFIANÇA NA CLASSIFICAÇÃO; A4=ELEVADA CONFIANÇA NA CLASSIFICAÇÃO; A5=MUITO ELEVADA CONFIANÇA NA CLASSIFICAÇÃO. ....	34
TABELA 14. CONTRIBUIÇÕES PONDERADAS DE CADA CATEGORIA DE CONF PARA O VALOR DE EXACTIDÃO $P_{CONF}$ . O VALOR DO PESO PARA CADA CATEGORIA DE CONF ESTÁ LISTADO ENTRE PARÊNTESES POR BAIXO DE CADA CATEGORIA. A COLUNA ‘ACORDO’, INDICA A CONTRIBUIÇÃO PARA $P_{CONF}$ , DE TODAS AS OBSERVAÇÕES QUE CORRESPONDEM A ACERTOS ENTRE A CLASSE DE REFERÊNCIA	

E MAPA (I.E. CATEGORIAS *A1*, *A2*, *A3*, *A4* E *A5*).  $P_{CONF}$  MAX, REPRESENTA O VALOR MÁXIMO POSSÍVEL DE  $P_{CONF}$ , SE TODAS AS OBSERVAÇÕES EM CADA CATEGORIA DE CONF, REPRESENTAREM ACORDO ENTRE A CLASSE DO MAPA E A CLASSE DE REFERÊNCIA PRINCIPAL (L1). OS ACRÓNIMOS DAS CATEGORIAS CONF SÃO: *E3*=ERRO MUITO ELEVADO; *E2*=ERRO ELEVADO; *E1*=ERRO ACEITÁVEL; *A1*=MUITO REDUZIDA CONFIANÇA NA CLASSIFICAÇÃO; *A2*=REDUZIDA CONFIANÇA NA CLASSIFICAÇÃO; *A3*=ACEITÁVEL CONFIANÇA NA CLASSIFICAÇÃO; *A4*=ELEVADA CONFIANÇA NA CLASSIFICAÇÃO; *A5*=MUITO ELEVADA CONFIANÇA NA CLASSIFICAÇÃO. .... 36

TABELA 15. PROPORÇÃO DE OBSERVAÇÕES EM CADA CATEGORIA LOC POR CLASSE DE OCUPAÇÃO DO SOLO, E AS MEDIDAS DE EXACTIDÃO TEMÁTICA MAX, RIGHT E  $P_{LOC}$ . AS MEDIDAS DE EXACTIDÃO APRESENTADAS, SÃO ANÁLOGAS À EXACTIDÃO DO UTILIZADOR, PORQUE O DIVISOR É BASEADO NO NÚMERO DE OBSERVAÇÕES PRESENTES NO MAPA PARA CADA CLASSE DE OCUPAÇÃO DO SOLO. OS ACRÓNIMOS DAS CATEGORIAS LOC SÃO: *E3*=ERRO MUITO ELEVADO; *E2*=ERRO ELEVADO; *E1*=ERRO ACEITÁVEL; *A1*=MUITO REDUZIDA CONFIANÇA NA CLASSIFICAÇÃO; *A2*=REDUZIDA CONFIANÇA NA CLASSIFICAÇÃO; *A3*=RAZOÁVEL CONFIANÇA NA CLASSIFICAÇÃO; *A4*=ACEITÁVEL CONFIANÇA NA CLASSIFICAÇÃO; *A5*=ELEVADA CONFIANÇA NA CLASSIFICAÇÃO; *A6*=ELEVADA CONFIANÇA NA CLASSIFICAÇÃO. 38

TABELA 16. CONTRIBUIÇÕES PONDERADAS DE CADA CATEGORIA DE LOC PARA O VALOR DE EXACTIDÃO  $P_{LOC}$ . O VALOR DO PESO PARA CADA CATEGORIA DE LOC ESTÁ LISTADO ENTRE PARÊNTESES POR BAIXO DE CADA CATEGORIA. A COLUNA ‘ACORDO’, INDICA A CONTRIBUIÇÃO PARA  $P_{LOC}$ , DE TODAS AS OBSERVAÇÕES QUE CORRESPONDEM A ACERTOS ENTRE A CLASSE DE REFERÊNCIA E MAPA (I.E. CATEGORIAS *A1*, *A2*, *A3*, *A4*, *A5* E *A6*). OS ACRÓNIMOS DAS CATEGORIAS LOC SÃO: *E3*=ERRO MUITO ELEVADO; *E2*=ERRO ELEVADO; *E1*=ERRO ACEITÁVEL; *A1*=MUITO REDUZIDA CONFIANÇA NA CLASSIFICAÇÃO; *A2*=REDUZIDA CONFIANÇA NA CLASSIFICAÇÃO; *A3*=RAZOÁVEL CONFIANÇA NA CLASSIFICAÇÃO; *A4*=ACEITÁVEL CONFIANÇA NA CLASSIFICAÇÃO; *A5*=ELEVADA CONFIANÇA NA CLASSIFICAÇÃO; *A6*=MUITO ELEVADA CONFIANÇA NA CLASSIFICAÇÃO..... 40

TABELA 17. CONTRIBUIÇÕES PONDERADAS DE CADA CATEGORIA DE CONF, DEFINIDAS POR UM UTILIZADOR PARA O VALOR DE EXACTIDÃO  $P_{CONF,USER}$ . O VALOR DO PESO PARA CADA CATEGORIA DE CONF ESTÁ LISTADO ENTRE PARÊNTESES POR BAIXO DE CADA CATEGORIA. OS ACRÓNIMOS DAS CATEGORIAS CONF SÃO: A3=ACEITÁVEL CONFIANÇA NA CLASSIFICAÇÃO; A4=ELEVADA CONFIANÇA NA CLASSIFICAÇÃO; A5=MUITO ELEVADA CONFIANÇA NA CLASSIFICAÇÃO..... 41

TABELA 18. CONTRIBUIÇÕES PONDERADAS DE CADA CATEGORIA DE LOC, DEFINIDAS POR UM UTILIZADOR PARA O VALOR DE EXACTIDÃO  $P_{LOC,USER}$ . O VALOR DO PESO PARA CADA CATEGORIA DE LOC ESTÁ LISTADO ENTRE PARÊNTESES POR BAIXO DE CADA CATEGORIA. OS ACRÓNIMOS DAS CATEGORIAS LOC SÃO: A4=ACEITÁVEL CONFIANÇA NA CLASSIFICAÇÃO; A5=ELEVADA CONFIANÇA NA CLASSIFICAÇÃO; A6=MUITO ELEVADA CONFIANÇA NA CLASSIFICAÇÃO. .... 42

## ÍNDICE DE FIGURAS

FIGURA 1. REPRESENTAÇÃO MATEMÁTICA DE UMA MATRIZ DE CONFUSÃO (ADAPTADO DE CONGALTON E GREEN, 1999). .....	2
FIGURA 2. EXEMPLO DE UMA UNIDADE AMOSTRAL (BLOCO DE 3X3 PIXÉIS).....	16
FIGURA 3. EXEMPLOS PARA O ÍNDICE DE CONFIANÇA DE INTERPRETAÇÃO (ICR) (EM CIMA) E PARA O ÍNDICE DE CONFIANÇA DE LOCALIZAÇÃO (LCR). .....	16
FIGURA 4. EXEMPLO DE CADA UMA DAS CLASSES DE OCUPAÇÃO DO SOLO DA NOMENCLATURA UTILIZADA NO ESTUDO PROPOSTO. ....	27
FIGURA 5. MAPA DE OCUPAÇÃO DO SOLO PRODUZIDO A PARTIR DO COMPÓSITO BIMENSAL DERIVADO DO SENSOR MERIS (ADAPTADO DE CARRÃO <i>ET AL.</i> 2009). .....	28
FIGURA 6. DISTRIBUIÇÃO ESPACIAL DA AMOSTRA DE REFERÊNCIA PELO TERRITÓRIO DE PORTUGAL CONTINENTAL. ....	30

# 1. INTRODUÇÃO

A ocupação do solo é uma variável fundamental para a compreensão da interacção e dos impactes, que as actividades humanas impõem ao meio biofísico em que nos inserimos. Em nenhuma outra altura da história da civilização humana, o Homem imprimiu alterações tão profundas na ocupação do solo, sendo assim necessário perceber que influências poderão ter estas alterações no nosso estilo de vida, como também na nossa própria sobrevivência enquanto Humanidade. Segundo Lesschen *et al.* (2005), as alterações na ocupação do solo são o resultado da interacção entre factores socioeconómicos, institucionais e ambientais. As alterações na ocupação do solo são referidas como a variável mais importante na alteração dos sistemas ecológicos a um nível global (Vitousek 1994); como factor determinante na alteração do sistema climático (Brovkin *et al.* 2004); como a variável com maior impacte na biodiversidade nos próximos 100 anos (Chapin *et al.* 2000).

Importa assim perceber o significado e importância que a ocupação do solo e as suas alterações têm no sistema global. No entanto, para uma monitorização da ocupação do solo a uma escala global, em curtos espaços de tempo e com custos que as empresas que produzem mapas de ocupação do solo possam comportar, é necessário recorrer a dados obtidos através de detecção remota, mais concretamente, a imagens recolhidas por satélites. De facto, Foody (2002) refere que as imagens de satélite são uma fonte bastante atractiva na produção de mapas de ocupação do solo, na medida em que permitem uma representação espacialmente contínua, a várias escalas espaciais e temporais da superfície da Terra. Por estes motivos, os dados obtidos através de satélite são nos dias de hoje uma fonte crescente de informação para produção de mapas de ocupação do solo (Caetano *et al.* 2006).

Apesar da produção de mapas de ocupação do solo ser de extrema importância é também importante verificar como estes mapas representam a realidade. Este é um aspecto muitas vezes renegado para segundo plano, que pode ter fortes implicações nas decisões tomadas pelos utilizadores finais destes mapas. Se são tomadas decisões baseadas nesta informação, então a melhor ou pior qualidade desta informação influenciará também a qualidade das decisões tomadas. Por estes factos, os mapas de

ocupação do solo, antes de poderem ser utilizados para investigação científica e processos de tomada de decisão, devem ser sujeitos a uma rigorosa avaliação estatística da sua exactidão (Stehman e Czaplewski 1998).

### 1.1 Avaliação da exactidão temática de mapas de ocupação do solo

Tradicionalmente, a avaliação da exactidão temática de mapas de ocupação do solo é realizada através da comparação dos mapas produzidos com uma base de dados de referência, sendo esta comparação feita para um conjunto de pontos ou zonas amostra e representada numa matriz de confusão (Story e Congalton 1986). Geralmente a informação de referência é representada nas colunas da matriz, sendo comparada com a informação do mapa representada nas linhas da matriz. Os valores na diagonal da matriz representam o acordo entre ambos os conjuntos de dados, ou seja, entre a classificação do mapa e os dados de referência. A matriz de confusão, por permitir uma análise individual da exactidão temática por classe, é considerada um método bastante válido para representar a exactidão de mapas (Congalton e Green 1999). Na Figura 1 é apresentada uma representação matemática de uma matriz de confusão.

		j = colunas (referência)			Total da linha $n_{i+}$
		1	2	k	
i = linhas (classificação)	1	$n_{11}$	$n_{12}$	$n_{1k}$	$n_{1+}$
	2	$n_{21}$	$n_{22}$	$n_{2k}$	$n_{2+}$
	k	$n_{k1}$	$n_{k2}$	$n_{kk}$	$n_{k+}$
	Total da coluna $n_{+j}$	$n_{+1}$	$n_{+2}$	$n_{+k}$	$n$

Figura 1. Representação matemática de uma matriz de confusão (Adaptado de Congalton e Green, 1999).

A exactidão global do mapa é calculada através da divisão do somatório das unidades amostrais da diagonal da matriz, pelo número total de unidades amostrais da amostra (Story e Congalton 1986). Este índice é o mais comum e reflecte a proporção de área correctamente classificada (Stehman 1997), em que  $i$  representa as linhas da matriz,  $n_{ii}$  representa o número de unidades amostrais correctamente classificadas para cada classe de ocupação do solo  $k$ , e  $n$  a dimensão da amostra, sendo dado por



$$\text{Exactidão global} = \sum_{i=1}^k \frac{n_{ii}}{n} \quad (1)$$

Story e Congalton (1986) referem outro método de cálculo da exactidão temática, no qual o número de unidades amostrais correctamente classificadas ( $n_{kk}$ ) é dividido pelo número total das unidades amostrais de referência (total da coluna), ou seja, por  $n_{+k}$ . Esta proporção representa a probabilidade das unidades amostrais de referência serem bem classificadas, ou seja, estão a ser medidos os erros de omissão. Os erros de omissão surgem quando é excluída uma área no mapa que pertence a uma determinada classe de ocupação do solo (Congalton e Green 1999). Este valor é denominado de exactidão do produtor e indica ao produtor do mapa como pode estar correctamente classificada uma determinada área no terreno. O índice de exactidão do produtor para uma classe de ocupação do solo  $j$ , indica qual a probabilidade de uma classe de ocupação do solo  $j$  nos dados de referência ser classificada como  $j$ , no mapa produzido (Stehman 1997), sendo dado por

$$\text{Exactidão do produtor} = \frac{n_{jj}}{n_{+j}} \quad (2)$$

Por outro lado, quando é dividido o número de unidades amostrais correctamente classificadas ( $n_{kk}$ ) de uma determinada classe de ocupação do solo, pelo número total de unidades amostrais dessa classe representada no mapa ( $n_{k+}$ ), este valor indica a probabilidade de uma determinada classe de ocupação do solo no mapa representar essa mesma classe na realidade. Neste caso estão a ser medidos os erros de comissão, que são definidos como a inclusão de uma área no mapa, numa classe de ocupação do solo na qual essa área não poderia ter sido incluída (Congalton e Green 1999). Este valor é definido como exactidão do utilizador e indica como o mapa representa a realidade. O índice de exactidão do utilizador para uma classe de ocupação do solo  $i$ , indica a probabilidade de uma área classificada como  $i$  no mapa produzido, ser classificado como  $i$  nos dados de referência (Stehman 1997), sendo dado por

$$\text{Exactidão do utilizador} = \frac{n_{ii}}{n_{i+}} \quad (3)$$

## **1.2 Limitações da abordagem tradicional de avaliação da exactidão temática de mapas de ocupação do solo**

Nos métodos tradicionalmente utilizados na avaliação da exactidão temática de mapas de ocupação do solo, são assumidos vários aspectos que limitam a compreensão da exactidão dos mapas de ocupação do solo. Segundo Gopal e Woodcock (1994) e Woodcock e Gopal (2000), é assumido que a cada ponto ou zona amostra do mapa apenas pode ser atribuída uma e só uma classe de ocupação do solo, sendo assumido que esta atribuição é feita com toda a certeza. Para a elaboração de uma base de dados de referência o técnico tem que, em visitas de campo ou através de foto-interpretação de imagens aéreas de elevada resolução espacial, escolher uma classe para cada localização, mesmo que exista mais do que uma classe possível nessa mesma localização (Gopal e Woodcock 1994). Este aspecto verifica-se principalmente em zonas de transição entre classes de ocupação do solo e por esta apresentar um *continuum* natural, pois raramente a ocupação do solo apresenta transições abruptas entre classes. Por este motivo, os mapas de ocupação do solo são vagos e possuem uma forte componente de incerteza, não sendo expectável que duas pessoas cheguem a classificações iguais (Goodchild 2003).

Outro aspecto que é assumido na abordagem tradicional de avaliação da exactidão temática de mapas de ocupação do solo é o facto de ser considerado um co-registo perfeito entre os mapas produzidos e as bases de dados de referência. No entanto este aspecto não se verifica (nomeadamente em mapas de média/reduzida resolução espacial), e os erros dos mapas produzidos podem não ser de origem temática mas sim devido ao não co-registo perfeito entre ambos os conjuntos de dados. Hagen (2003) designa este aspecto de “incerteza posicional”, para expressar o facto de uma localização específica num mapa não ser sempre tão exacta como aparenta. Uma determinada classe posicionada num mapa numa determinada localização pode efectivamente estar presente na proximidade dessa localização, sendo o erro provocado por dificuldades de posicionamento. De facto, não é expectável um co-registo perfeito entre uma localização de um mapa e uma observação de uma base de dados de referência, podendo este facto enviesar a aparente exactidão temática de um

mapa de ocupação do solo. Sem este co-registo perfeito a matriz de confusão pode conter erros temáticos e de posição, o que dificulta a correcta interpretação das métricas de exactidão derivadas da matriz de confusão (Foody 2002). Sendo assim, a assunção de que as bases de dados de referência são isentas de erros, não se verifica (Congalton e Green 1999), podendo este aspecto levar a enviesamento dos índices de exactidão temática que são disponibilizados aos utilizadores finais de mapas de ocupação do solo.

Na abordagem tradicional de avaliação da exactidão temática de mapas de ocupação do solo, é também assumido que os erros dos mapas possuem todos a mesma importância e que estão limitados ao registo da não concordância entre as classes do mapa e as classes da base de dados de referência. A informação acerca da magnitude ou gravidade dos erros não é utilizada. Informação mais detalhada sobre os erros e a incerteza poderá ajudar os utilizadores a verificarem se os mapas satisfazem ou não as suas necessidades (Gopal e Woodcock 1994). Assim sendo, os utilizadores necessitam que lhes seja fornecida informação mais completa e interpretável sobre o mapa que estão a utilizar (Aronoff 1982). Vários autores referem que um importante tópico na avaliação da exactidão temática de mapas de ocupação do solo é a avaliação da magnitude e gravidade dos erros, e que em abordagens anteriores os erros têm sido tratados de forma semelhante (Gómez *et al.* 2008, Foody 2002).

### **1.3 Incerteza na elaboração de bases de dados de referência**

A informação de referência utilizada na avaliação da exactidão temática de mapas de ocupação do solo é obtida a partir de uma amostra seleccionada aleatoriamente sobre a área geográfica em estudo. Tradicionalmente é assumido que em cada uma das unidades amostrais da base de dados de referência apenas pode existir uma classe de ocupação do solo, que será a que melhor representa essa mesma área geográfica. No entanto, devido à heterogeneidade da paisagem e ao *continuum* natural da ocupação do solo, a escolha da classe de ocupação que melhor caracteriza uma determinada área não é tarefa fácil. Assim, a escolha de apenas uma classe pode influenciar a avaliação da exactidão temática de um mapa de ocupação do solo.

Para evitar este problema, aquando da elaboração de bases de dados de referência, pode ser recolhida para cada unidade amostral mais do que uma classe de ocupação do solo (Woodcock *et al.* 1996, Edwards *et al.* 1998, Zhu *et al.* 2000, Wickham *et al.* 2004), sendo a abordagem mais utilizada, a recolha de uma classe de ocupação do solo principal e outra secundária em cada uma das observações da base de dados de referência. Outra informação que pode ser recolhida é um índice de confiança de interpretação (*Interpretation confidence rating* (ICR)), o qual reflecte a confiança com que o foto-intérprete atribui uma determinada classe de ocupação do solo em cada uma das observações (Muchoney *et al.* 1999, Zhu *et al.* 2000, Wickham *et al.* 2004). Deste modo, a incerteza presente nas bases de dados de referência é representada de duas formas distintas: através da recolha de duas classes de ocupação do solo rígidas e da recolha de uma variável qualitativa, que reflecte a incerteza na recolha das classes de ocupação do solo de referência. A escala de confiança pode variar de estudo para estudo, sendo definida pela percepção do analista. Em estudos anteriores, foram utilizadas várias escalas; de 1 a 9 (Muchoney *et al.* 1999); de 1 a 4 em que 1 é definido como ‘*A informação sobre a ocupação/uso do solo é muito difícil de interpretar*’ e 4 ‘*A interpretação é absolutamente correcta*’ (Zhu *et al.* 2000); escala de 1 a 4 em que 1 é ‘*Não confiante*’ e 4 sendo ‘*Muito confiante*’ (Wickham *et al.* 2004); e uma medida subjectiva de certeza atribuída às classes de ocupação do solo em cada uma das observações, aquando de visitas de campo (Edwards *et al.* 1998).

Segundo Lanter e Veregin (1992), a incerteza posicional e temática não são independentes. Neste sentido, e com o intuito de incluir a incerteza posicional de cada uma das observações de bases de dados de referência, vários autores sugerem ainda a recolha de um índice baseado na localização (*Location confidence rating* (LCR)) (Zhu *et al.* 2000; Stehman *et al.* 2003). Este índice traduz a percepção do técnico relativamente às características da paisagem em cada uma das observações bem como na sua vizinhança (geralmente é observado um bloco de 3x3 pixéis em torno do pixel central), sendo este índice de carácter qualitativo. A escala geralmente utilizada possui três níveis em que 1 é ‘*No limite entre classes de ocupação do solo*’; 2 ‘*Ocupação do solo homogénea*’ e 3 ‘*Ocupação do solo heterogénea*’ (Yang *et al.* 2000; Yang *et al.* 2001; Zhu *et al.* 2000).

Para além dos métodos de inclusão da incerteza na elaboração de bases de dados de referência acima mencionados, Gopal e Woodcock (1994) desenvolveram várias funções baseadas na teoria dos conjuntos *fuzzy* que permitem incluir a incerteza na elaboração de bases de dados de referência na avaliação da exactidão temática de mapas de ocupação do solo. As funções MAX e RIGHT permitem o cálculo da frequência de concordância e não concordância entre a classificação do mapa e a informação de referência; e a função DIFFERENCE que permite avaliar a magnitude e gravidade do erro e analisar os erros do mapa de forma diferenciada. Para a implementação destas funções os autores desenvolveram uma escala linguística, para que os técnicos pudessem descrever a sua percepção da ocupação do solo em cada uma das unidades amostrais da base de dados de referência através de visitas de campo. Na Tabela 1 são apresentados os valores da escala linguística e a respectiva descrição utilizada neste estudo.

Valores da escala linguística	Descrição
1	<i>Classe absolutamente errada</i> : classe inaceitável; absolutamente errada;
2	<i>Percebe-se, mas errada</i> : não é uma boa classe. Existe algum factor no local que faz com que se perceba a classe, mas existe claramente uma melhor classe. Esta classe pode ser um problema para os utilizadores do mapa. Não está correcta;
3	<i>Classe aceitável ou razoável</i> : pode não ser a melhor classe possível, mas é aceitável; esta classe não coloca qualquer problema ao utilizador se vier expressa no mapa. Correcta;
4	<i>Boa classe</i> : classe muito correcta;
5	<i>Classe absolutamente correcta</i> : sem dúvida acerca da compatibilidade. Classe perfeita.

Tabela 1. Escala linguística e respectiva descrição utilizada no trabalho de Gopal e Woodcock (1994) (Adaptado de Gopal e Woodcock 1994).

Com esta abordagem o técnico não está limitado apenas a uma resposta correcta ou errada em cada uma das unidades amostrais, ao contrário dos métodos tradicionalmente utilizados na avaliação da exactidão temática de mapas de ocupação do solo.

*- Frequência de concordância e não concordância entre a classificação do mapa e a informação de referência*

Nesta fase da avaliação da exactidão temática pretende-se avaliar a exactidão do mapa produzido, em termos da frequência de concordância entre a classificação do mapa ( $M$ ) e a informação da amostra de referência ( $A$ ). Este método divide o problema tradicional “qual a exactidão do mapa”, em duas questões mais precisas:

- 1) *Com que frequência é a classe classificada no mapa a melhor escolha para representar a realidade?*
- 2) *Com que frequência é a classe classificada no mapa aceitável para representar a realidade?*

Utilizando a notação utilizada na obra de Gopal e Woodcock (1994), ou seja, considerando  $\sigma$  uma função *booleana* na qual se obtém um resultado de 0 ou 1, segundo a pertença da unidade amostral  $x$  à classe  $C$ , relativamente a  $A$ . A avaliação da pertença da unidade amostral  $x$  do conjunto  $S$  ( $x \in S$ ) à classe  $C$  do mapa, depende do critério utilizado para definir  $\sigma$ .  $\sigma(x, C) = 1$  se  $x$  pertencer a  $C$ , e  $\sigma(x, C) = 0$  se  $x$  não pertencer a  $C$ . Considerando agora duas diferentes definições de  $\sigma$ . Se  $\sigma(x, \aleph(x)) = 1$ , então pode-se dizer que existe concordância entre os dados do mapa produzido e a avaliação do especialista na unidade amostral  $x$  na realidade, e se  $\sigma(x, \aleph(x)) = 0$ , então não existe concordância entre os dois conjuntos de dados. Para cada classe do mapa  $C \in C$ , calculam-se duas quantidades:

$$\omega_C = |\{x \in S \mid \aleph(x) = C \text{ e } \sigma(x, C) = 1\}|$$

$$\overline{\omega}_C = |\{x \in S \mid \aleph(x) = C \text{ e } \sigma(x, C) = 0\}|$$

Ou seja, conta-se o número de unidades amostrais  $x$  com  $\aleph(x) = C$  em termos de concordância ( $\omega_C$ ) e não concordância ( $\overline{\omega}_C$ ). A concordância e não concordância pode ser definida através de duas funções:

- 1) Função MAX

A função MAX pode ser definida como:

$$\text{MAX}(x, C) = \begin{cases} 1 & \text{se } \mu_C(x) \geq \mu_{C'}(x) \text{ para todos os } C' \in C, \\ 0 & \text{se } \mu_C(x) \leq \mu_{C'}(x) \text{ para todos os } C' \in C \end{cases}$$

$\text{MAX}(x, C) = 1$  se o valor da função de pertença  $x$  na classe  $C$  ( $\mu_C(x)$ ), é máximo ao longo de todas as classes do mapa ( $\mu_{C'}(x)$ ).

## 2) Função RIGHT

A função RIGHT é definida tendo como ponto inicial, um limiar predefinido  $t$  (no trabalho desenvolvido por Gopal e Woodcock (1994) este limiar é  $\geq 3$  na escala linguística apresentada), em que  $t$  é um valor da escala linguística ou um valor específico da função de pertença. Uma unidade amostral  $x$  “pertença” a uma classe  $C$  se a sua função de pertença for  $\mu_C(x) \geq t$ . A função RIGHT pode ser definida como:

$$\text{RIGHT}(x, C) = \begin{cases} 1 & \text{se } \mu_C(x) \geq t \\ 0 & \text{se } \mu_C(x) < t \end{cases}$$

Normalmente a função MAX é mais conservadora do que a função RIGHT, mas este facto nem sempre acontece. Este aspecto pode acontecer devido à possibilidade de existirem unidades amostrais em que a classificação do mapa não possua classes que coincidam com a função RIGHT, mas que mesmo assim tenham um valor elevado da escala linguística (Gopal e Woodcock, 1994). O mesmo autor refere ainda que na elaboração de matrizes de confusão, geralmente são utilizadas as funções RIGHT e MAX. A função MAX é aplicada quando a matriz de confusão é elaborada a partir de unidades amostrais “cegas” (quando a classificação do mapa não é conhecida, na fase de elaboração da base de dados de referência). Por outro lado a função RIGHT é utilizada quando existe o conhecimento da classificação do mapa durante a avaliação das unidades amostrais. Muitas vezes o conhecimento da classificação do mapa durante a sua avaliação não é fornecido numa matriz de confusão. Separando os resultados das funções RIGHT e MAX, permite que não exista confusão no aspecto anteriormente referido. Os dados das funções RIGHT e MAX quando obtidos em conjunto, são mais úteis do que quando obtidos em separado.

Relativamente ao tratamento dos erros do mapa de forma diferenciada, na obra pioneira de Gopal e Woodcock (1994), os autores elaboraram um operador que permite avaliar a magnitude e gravidade do erro que se denomina de DIFFERENCE.

Com o intuito de avaliar a magnitude dos erros, consideremos a função  $\Delta(x) \rightarrow \mathbb{Z}$ . Para uma determinada unidade amostral  $x$ ,  $\Delta(x)$  mede a diferença entre o resultado obtido pela classe  $\kappa(x)$  em  $x$  presente no mapa e o valor mais elevado atribuído a  $x$  entre todas as outras categorias de  $C$  relativamente à informação dos dados de referência. A função acima descrita é denominada de DIFFERENCE e pode ser definida como:

$$\text{DIFFERENCE}(\Delta(x)) = \mu_{\kappa(x)}(x) - \max \mu_C(x)$$

Como a escala linguística utilizada neste estudo varia entre 1 e 5,  $\Delta(X)$  está compreendido entre -4 e 4. Os valores negativos deste operador representam, o não acordo entre a classe de referência e a classe presente no mapa (valores mais negativos, representam erros mais graves), e os valores positivos representam o acordo entre ambos os conjuntos de dados (valores mais elevados correspondem a valores de exactidão mais elevados).

Na situação ideal (na qual a classe do mapa tem valor linguístico 5 e todas as outras classes são consideradas completamente erradas (valor linguístico 1), a função DIFFERENCE apresenta valor 4. Para todas as unidades amostrais em que existe concordância entre a classificação do mapa e a informação de referência, se for utilizada a função MAX, terão valores da função DIFFERENCE superiores ou iguais a 0.

Por outro lado, as amostras nas quais se obtêm valores negativos correspondem a unidades amostrais em que a classificação do mapa e a informação de referência são incompatíveis. Uma amostra incompatível com um valor de DIFFERENCE de -1, corresponde a um caso em que se atribuiu apenas um valor abaixo na unidade amostral de referência, em relação ao possível valor máximo atribuível a essa mesma unidade amostral. Pelo contrário se a classe representada no mapa possui valor



linguístico 1 e alguma das outras classes da referência possui valor linguístico 5, então DIFFERENCE apresenta um valor de -4 (erro muito grave).

Gopal e Woodcock (1994) salientam o facto de que o uso desta função está dependente da forma como os valores da escala linguística são atribuídos. No exemplo apresentado pelos autores, a escala linguística foi convertida em valores discretos através de uma função linear simples, mas no entanto existem muito mais possibilidades.

A utilização dos valores obtidos a partir da função DIFFERENCE, em conjunto com as classes do mapa, pode fornecer índices muito úteis. Gopal e Woodcock (1994) referem que pode ser útil a atribuição de pesos, para acentuar os efeitos dos valores negativos, com o intuito de produzir um índice da magnitude do erro para cada classe.

No entanto o operador DIFFERENCE, não providencia uma medida de exactidão temática rígida que tenha em conta a incerteza inerente à elaboração de bases de dados de referência. A importância do reporte ao utilizador final de mapas de ocupação, de uma medida de exactidão temática que inclua a incerteza, para além das métricas tradicionais, é referida por Muller *et al.* (1998). Para além deste aspecto, na abordagem proposta por Gopal e Woodcock (1994), os erros são tratados de forma semelhante e os autores sugerem a utilização de um esquema de pesos, que permita o cálculo e tratamento dos erros de forma diferenciada. Neste sentido, no trabalho desenvolvido por Gómez *et al.* (2008) é utilizado um esquema de pesos, na elaboração de vários índices de exactidão temática, baseados na dissimilaridade entre a classificação do mapa e a classificação de referência. No entanto a determinação do esquema de pesos a utilizar é bastante subjectiva, estando bastante dependente dos requisitos do utilizador final (Sarmiento *et al.* 2009a).

#### **1.4 Motivações**

Previamente foram realizados estudos preliminares no Grupo de Detecção Remota (GDR) do Instituto Geográfico Português (IGP) (Sarmiento 2007; Sarmiento *et al.* 2008a), nos quais se iniciou o desenvolvimento de uma metodologia de avaliação da

exactidão temática de mapas de ocupação do solo que permitiu integrar a incerteza na elaboração de bases de dados de referência na avaliação da exactidão temática de mapas de ocupação do solo, como também avaliar os erros de forma diferenciada. Esta dissertação de mestrado é assim a continuação destes estudos prévios, dos quais resultaram três publicações científicas desenvolvidas durante a parte curricular do Mestrado em Ciência & Sistemas de Informação Geográfica do Instituto Superior de Estatística e Gestão de Informação (ISEGI) da Universidade Nova de Lisboa (UNL) (Sarmiento *et al.* 2008b; Sarmiento *et al.* 2009a; Sarmiento *et al.* 2009b), fazendo os mesmos parte integrante desta tese.

Esta dissertação de mestrado enquadra-se em linhas de investigação definidas como prioritárias pela *European Space Agency* (ESA), através de iniciativas como o *Global Observation for Forest and Land Cover Dynamics* (GOFC-GOLD), na qual é promovida também a colaboração e cooperação com o *Committee on Earth Observation Satellites – Working Group on Calibration & Validation* (CEOS-WGCV) inserido na *National Aeronautics and Space Administration* (NASA). Algumas das actividades prioritárias de ambas estas iniciativas, é o de promover a cooperação internacional no desenvolvimento de metodologias de calibração/validação de dados de observação da Terra e desenvolver um programa operacional de validação de mapas globais de ocupação do solo. Neste sentido, as linhas de investigação que se pretendem desenvolver e os objectivos que se pretende atingir, vêm no sentido do melhoramento das metodologias de avaliação da exactidão temática de mapas de ocupação do solo e de tópicos nesta mesma área, que necessitam de mais investigação. É de salientar que os aspectos acima mencionados vêm de encontro às preocupações e linhas de investigação que necessitam de desenvolvimento, referidas no manual de Recomendações para a Avaliação da Exactidão Temática de Mapas Globais de Ocupação do Solo, elaborado pelo CEOS-WGCV (Strahler *et al.* 2006).

## **1.5 Objectivos**

Pretende-se com a dissertação de mestrado proposta, desenvolver uma metodologia que permita incorporar a incerteza na atribuição das classes de ocupação do solo de referência, a incerteza de localização e a incerteza temática na elaboração de bases de

dados de referência, de modo a serem elaborados índices de exactidão temática rígidos que reflectam os três tipos de incerteza acima mencionados. Os índices de exactidão temática propostos visam complementar as métricas tradicionalmente utilizadas (i.e. exactidão do utilizador, do produtor e global), permitindo assim o reporte aos utilizadores finais, dos impactes que a incerteza na elaboração de bases de dados de referência podem ter na aparente exactidão temática de mapas de ocupação do solo.

De forma a ilustrar a metodologia proposta, recorreu-se a um caso de estudo onde se avaliou a qualidade temática de um mapa de ocupação do solo de Portugal Continental derivado de imagens MERIS (*Medium Resolution Imaging Spectrometer Instrument*).

## **1.6 Estrutura da dissertação**

A dissertação está dividida em 4 capítulos, em que no presente capítulo se faz uma breve abordagem aos métodos tradicionalmente utilizados na avaliação da exactidão temática de mapas de ocupação do solo e das suas respectivas limitações. Para além disso são apresentados de forma sumária, algumas metodologias que permitem estudar a incerteza na elaboração de bases de dados de referência na avaliação da exactidão temática de mapas de ocupação do solo.

No capítulo 2, faz-se uma descrição detalhada da metodologia desenvolvida, para incluir a incerteza na elaboração de bases de dados de referência, de forma a serem elaborados índices de exactidão temática rígidos.

Posteriormente, é apresentado no capítulo 3 um caso de estudo no qual se testou a metodologia proposta, e no qual são apresentados os resultados obtidos.

Finalmente, no capítulo 4 são apresentadas as conclusões, como também as limitações e linhas de desenvolvimento possíveis.

## **2. METODOLOGIA**

Neste capítulo da dissertação é apresentada a metodologia que se pretende implementar para alcançar os objectivos propostos. Inicialmente, foi necessária a recolha da informação de referência que permitiu posteriormente o cálculo dos índices de exactidão temática tradicionais e a elaboração dos operadores CONF e LOC. Através de um esquema de pesos e da sua aplicação em ambos os operadores, foram calculadas as medidas de exactidão temática com inclusão da incerteza.

### **2.1 Base de dados de referência**

Antes de poder ser implementada a metodologia proposta de avaliação da exactidão temática de mapas de ocupação do solo, foi recolhida informação relativa à ocupação do solo em cada observação da amostra de referência. Mais concretamente, neste estudo foram recolhidas em cada observação duas classes de ocupação do solo (uma principal (L1) e uma secundária (L2)). A integração da incerteza na atribuição das classes de referência em cada observação foi realizada através da recolha de um índice de confiança na interpretação (ICR), com três níveis. Foi também recolhido um índice de confiança na localização (LCR), para reflectir a percepção do técnico relativamente às características da paisagem em cada uma das observações. Também este índice é constituído por três níveis. Tal como no trabalho desenvolvido por Zhu *et al.* (2000), as motivações para se recolherem duas classes de ocupação do solo deveu-se à possível ambiguidade e incerteza que existe entre classes de ocupação do solo, que na sua definição são bastante semelhantes (e.g. floresta de resinosas e floresta mista), considerando-se assim desta forma que ambas as classes de ocupação do solo podem ser consideradas correctas na descrição de uma determinada observação da amostra de referência. Os mesmos autores recolheram também informação relativa à confiança na atribuição das classes de ocupação do solo que reflecte a incerteza que houve durante o processo de foto-interpretação, e um índice de localização para reflectir possíveis problemas devido à existência de pixéis mistos. A escala de confiança na interpretação utilizada pelos autores contempla quatro níveis, enquanto que no estudo proposto foram utilizados três níveis. Para além disso, na metodologia proposta neste estudo os índices de localização foram

reordenados de forma a transmitir uma escala de homogeneidade da paisagem (i.e. LCR=3 corresponde a áreas em que a ocupação do solo é homogénea e LCR=1 corresponde a áreas em que a ocupação do solo é heterogénea), diferindo assim do trabalho desenvolvido por Zhu *et al.* (2000). Na Tabela 2 é apresentada a informação recolhida em cada uma das observações da base de dados de referência.

Classe de ocupação do solo principal (L1)		Classe de ocupação do solo considerada mais correcta pelo foto-intérprete.
Classe de ocupação do solo secundária (L2)		Classe de ocupação do solo que pode ser também considerada correcta pelo foto-intérprete.
Índice de confiança de interpretação (ICR)	ICR 3	Se pelo menos 81% da área for ocupada por uma classe de ocupação do solo.
	ICR 2	Se duas classes de ocupação do solo estão presentes e nenhuma delas ocupa mais de 81% da área.
	ICR 1	Se três ou mais classes de ocupação do solo estão presentes, nenhuma ocupando mais de 50 % da área.
Índice de confiança de localização (LCR)	LCR 3	Ocupação do solo homogénea no pixel central e sua vizinhança.
	LCR 2	No limite entre duas classes de ocupação do solo.
	LCR 1	Ocupação do solo heterogénea no pixel central e sua vizinhança.

Tabela 2. Informação recolhida em cada uma das observações da amostra de referência (Adaptado de Zhu *et al.* 2000).

Segundo Stehman e Czaplewski (1998) as unidades amostrais que podem ser utilizadas na avaliação da exactidão temática de mapas de ocupação do solo podem ser de dois tipos: 1) sob a forma de pontos; 2) sob a forma de áreas, em que neste último caso pode apresentar várias extensões e formas. Neste estudo as unidades amostrais utilizadas na elaboração da base de dados de referência consistem em blocos de 3x3 pixéis em torno do pixel central (300 X 300m), que corresponde à dimensão de um pixel de imagens MERIS. Na Figura 2 é apresentado um exemplo de uma unidade amostral utilizada neste estudo.

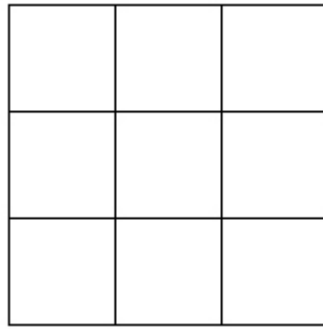


Figura 2. Exemplo de uma unidade amostral (bloco de 3x3 pixéis).

Na Figura 3 são apresentados exemplos dos índices de confiança de interpretação (ICR) e de localização (LCR).

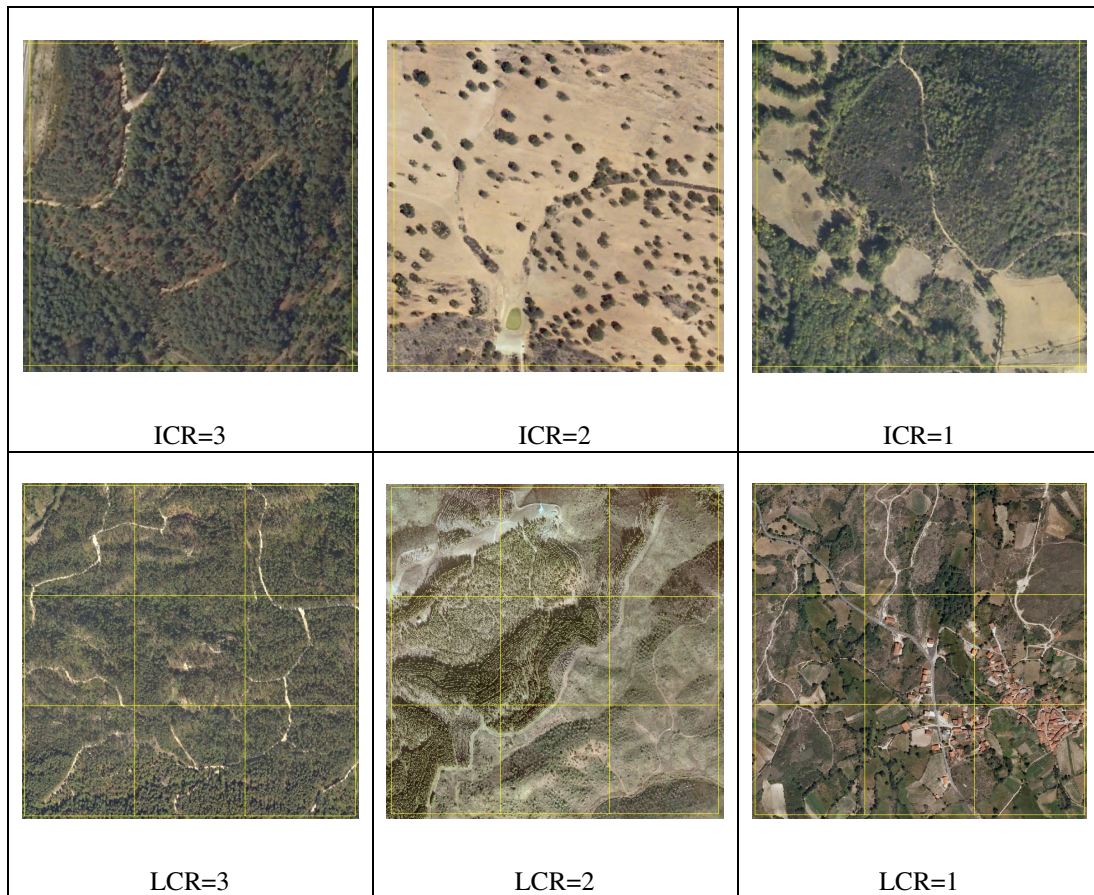


Figura 3. Exemplos para o índice de confiança de interpretação (ICR) (em cima) e para o índice de confiança de localização (LCR).

## 2.2 Avaliação da exactidão temática através dos métodos tradicionais

A possibilidade da recolha de mais do que uma classe de ocupação do solo de referência é o factor chave que permite a ligação com as medidas de exactidão MAX e RIGHT (Gopal e Woodcock 1994). Supondo que a informação de referência foi

recolhida tendo em conta a escala linguística proposta por Gopal e Woodcock (1994) apresentada na Tabela 1 e que não existem observações às quais foi atribuído um valor da escala linguística superior a 2 (*‘Percebe-se, mas errada’*) (i.e. em nenhuma das observações existem mais do que duas classes de ocupação do solo às quais foi atribuído um valor da escala linguística igual ou superior a 3, em que 3 é *‘Classe aceitável ou razoável’*, 4 é *‘Boa classe’* e 5 é *‘Classe absolutamente correcta’*). Neste sentido e considerando que não existem observações com um valor de escala linguística superior a 2, então o protocolo de recolha da informação de referência pode contemplar apenas a recolha de duas classes de ocupação do solo de referência (uma principal e outra alternativa). Laba *et al.* (2002) e Muller *et al.* (1998) reportaram respectivamente para as observações utilizadas nos seus estudos, que 94% e 100% das mesmas apresentavam valores da escala linguística de 1 e 2. Assim, para estas duas aplicações citadas na literatura, raramente existem observações em que o valor da escala linguística ultrapassa o valor 2.

Nos métodos tradicionais de avaliação da exactidão temática é apenas utilizada uma classe de ocupação do solo de referência em cada uma das observações. Esta classe de ocupação do solo pode ser vista como a classe de ocupação do solo considerada mais correcta pelo foto-intérprete (L1) presente Tabela 2. A exactidão global e por classe é calculada através do acordo entre a classe de ocupação do solo do mapa e L1 em cada uma das observações, sendo este método similar ao operador MAX definido por Gopal e Woodcock (1994). Como apenas é admitida uma classe de ocupação do solo de referência, para o cálculo da exactidão global e por classe, MAX pode ser visto como uma representação conservadora da exactidão temática. Quando estão disponíveis duas classes de ocupação do solo de referência (uma principal (L1) e uma alternativa (L2)), existe concordância entre ambos os conjuntos de dados, sempre que a classe de ocupação do solo presente no mapa coincida com L1 ou L2. Esta última abordagem tem em conta a possível ambiguidade na atribuição das classes de ocupação do solo em cada uma das observações, mas não tem em conta a incerteza inerente à atribuição das classes de ocupação do solo de referência e incerteza posicional, de uma base de dados de referência. Este método de cálculo é análogo ao operador RIGHT desenvolvido por Gopal e Woodcock (1994), sendo uma representação mais liberal da exactidão temática.

## 2.3 Incerteza da referência na avaliação da exactidão temática

Para incluir a incerteza na avaliação da exactidão temática, foram elaborados dois operadores. O operador CONF, o qual incorpora a incerteza na atribuição das classes de referência; e o operador LOC que incorpora a incerteza posicional. Ambos os operadores, reflectem a confiança que se poderá esperar da classificação do mapa. Tanto CONF como LOC são definidos pela combinação de respectivamente ICR e LCR, com o acordo ou desacordo entre as classes de ocupação do solo de referência (L1 e L2) e a classificação do mapa. Suponha-se que existe concordância entre a classificação do mapa e L1, e o foto-intérprete não teve qualquer dúvida na atribuição da classe de ocupação do solo de referência (ICR=3). Nesta situação pode-se afirmar que poderemos ter uma muito elevada confiança na classificação, categoria (A5) do operador CONF. Já nos casos em que não existe acordo entre ambas as classes de referência (L1 e L2) e a classificação do mapa, e estivermos perante uma zona em que não houve qualquer dúvida na interpretação da classe de referência (ICR=3), é considerado um erro muito elevado (E3). Estas situações podem ser devidas ao facto de o classificador não conseguir discriminar uma determinada classe de ocupação do solo. Na Tabela 3, são apresentadas as combinações de acertos e erros entre o mapa, L1, L2 e ICR.

Categorias CONF		L 1	L 2	ICR
<b>Acertos</b> (A)	Muito elevada confiança na classificação (A5)	√		3
	Elevada confiança na classificação (A4)	√		2
	Aceitável confiança na classificação (A3)		√	2
	Reduzida confiança na classificação (A2)	√		1
	Muito reduzida confiança na classificação (A1)		√	1
<b>Erros</b> (E)	Erro aceitável (E1)			1
	Erro elevado (E2)			2
	Erro muito elevado (E3)			3

Tabela 3. Categorias CONF agrupadas por acertos e erros entre a classificação do mapa e a classificação de referência. O acordo entre ambos os conjuntos de dados é especificado por (√).

Raciocinando desta forma, suponhamos que existe concordância entre a classificação do mapa e L1, e a unidade amostral situa-se numa zona homogénea (LCR=3). Nesta situação, pode-se afirmar que poderemos ter muito elevada confiança na classificação, categoria (A6) do operador LOC. Mesmo admitindo um deslocamento



da unidade amostral em relação à localização representada no mapa, como na ‘realidade’ estamos perante uma zona homogénea, pode-se afirmar com confiança que nessa mesma localização do mapa apenas existe na sua grande maioria, a classe de ocupação do solo representada no mapa. Por outro lado se não houver acordo entre L1 e a classe de ocupação do solo do mapa e estivermos perante uma zona homogénea, este erro é considerado muito elevado, categoria (E3) do operador LOC. O erro nestas situações pode indicar a incapacidade do classificador utilizado para produzir o mapa, de discriminar a classe de ocupação do solo presente na referência. Na Tabela 4 são apresentadas todas as combinações de erros e acertos entre a classificação do mapa, com L1, L2 e LCR, para o operador LOC. Como se pode verificar, o operador LOC apresenta seis categorias de acerto, enquanto que o operador CONF (Tabela 3) apresenta apenas cinco. Este facto deve-se à possibilidade de em unidades amostrais de referência que se situam em zonas homogéneas (LCR=3), de poder existir mais do que uma classe de ocupação do solo, havendo assim mais uma regra possível para o operador LOC, comparativamente ao operador CONF. Para o operador CONF, em unidades amostrais em que não existia qualquer dúvida na classe de ocupação do solo de referência (ICR=3), apenas foi considerada uma classe de ocupação do solo (L1=L2).

Categorias LOC		L 1	L 2	LCR
<b>Acertos</b> (A)	Muito elevada confiança na classificação (A6)	√		3
	Elevada confiança na classificação (A5)	√		2
	Aceitável confiança na classificação (A4)		√	3
	Razoável confiança na classificação (A3)		√	2
	Reduzida confiança na classificação (A2)	√		1
	Muito reduzida confiança na classificação (A1)		√	1
<b>Erros</b> (E)	Erro aceitável (E1)			1
	Erro elevado (E2)			2
	Erro muito elevado (E3)			3

Tabela 4. Categorias LOC agrupadas por acertos e erros entre a classificação do mapa e a classificação de referência. O acordo entre ambos os conjuntos de dados é especificado por (√).

## 2.4 Cálculo das medidas de exactidão temática $P_{CONF}$ e $P_{LOC}$

De modo a serem elaborados os índices de exactidão global e por classe, é atribuído a cada categoria dos operadores CONF e LOC um peso, que reflecte a total ou

parcial contribuição para a exactidão do mapa de cada uma das categorias de CONF (Tabela 5) e LOC (Tabela 6).

Exemplificando, observações que pertençam à categoria A5 da variável CONF e A6 da variável LOC, contribuem totalmente para a exactidão do mapa ( $W_d=1$ ), pois nestas situações pode-se ter total confiança de que a classificação do mapa é correcta. Por outro lado, às observações que pertencem à categoria A3 da variável CONF, e A4 e A3 da variável LOC, são atribuídas contribuições parciais de 0.5, visto nestas situações existir um acerto entre a classificação do mapa e a classe de ocupação do solo alternativa de referência (L2). Este aspecto indica que nessa localização existe outra classe de ocupação do solo, não sendo considerada a mesma confiança na classificação que no exemplo dado anteriormente.

As medidas de exactidão  $P_{CONF}$  (derivada de CONF) e  $P_{LOC}$  (derivada de LOC), são definidas da seguinte maneira. Suponhamos que existem  $M$  classes de ocupação do solo da nomenclatura adoptada com os códigos  $C_1$  a  $C_M$ , e cada observação da referência pertence a uma e apenas uma classe e a uma e apenas uma categoria  $D_{CONF} = \{A5, \dots, E3\}$  da variável CONF e a uma e apenas uma categoria  $D_{LOC} = \{A6, \dots, E3\}$  da variável LOC. Representando por  $p_{m,d}$  a proporção de área de cada uma das classes de ocupação do solo  $C_M$  na categoria  $d \in D_{CONF}$  da variável CONF e  $d \in D_{LOC}$  da variável LOC, e sendo  $W_{d,CONF}$  e  $W_{d,LOC}$  as contribuições de cada categoria  $d$  de CONF e LOC para a exactidão, a exactidão de  $P_{CONF,m}$  e  $P_{LOC,m}$  de cada classe  $C_M$  representada no mapa, é dada por

$$P_{CONF,m} = \sum_d p_{m,d} \times W_d \quad (4)$$

$$P_{LOC,m} = \sum_d p_{m,d} \times W_d \quad (5)$$

O somatório é calculado considerando todas as categorias  $D_{CONF}$  da variável CONF e  $D_{LOC}$  da variável LOC. Se  $p_{m,d}$  representa a proporção de área da classe  $C_M$  e da categoria  $d$  das variáveis CONF e LOC no mapa,  $P_{CONF,m}$  e  $P_{LOC,m}$  são similares à exactidão do utilizador, pois é condicionado pela área do mapa ocupada pela classe  $C_M$ . Se  $p_{m,d}$  representar a proporção de área da classe  $C_M$  na informação de referência

em cada categoria  $d$  das variáveis CONF e LOC, então  $P_{CONF,m}$  e  $P_{LOC,m}$  são similares à exactidão do produtor. O parâmetro que representa a exactidão global é dado por

$$P_{CONF} = \sum_d p_d \times W_d \quad (6)$$

$$P_{LOC} = \sum_d p_d \times W_d \quad (7)$$

em que  $p_d$  é a proporção de área na categoria  $d$  das variáveis CONF e LOC independentemente da classe de ocupação do solo.

Como só existem observações de referência para uma amostra,  $P_{CONF,m}$  e  $P_{LOC,m}$  podem apenas ser estimados utilizando os dados da amostra. Por exemplo, considerando uma amostragem aleatória simples ou uma amostragem aleatória estratificada e as classes do mapa como estratos,  $p_{m,d}$  poderá ser estimado considerando a proporção de área na amostra que pertence à classe  $C_M$  do mapa e à categoria  $d$  das variáveis CONF e LOC, obtendo-se neste caso uma medida semelhante à exactidão do utilizador.

#### 2.4.1 Elaboração do esquema de pesos

A atribuição de pesos a cada uma das categorias das variáveis CONF e LOC envolve alguma subjectividade, e a sua escolha poderá ser dependente do interesse do utilizador. Dificuldades semelhantes foram reportadas por outros autores (e.g. Gómez *et al.* 2008), não existindo uma metodologia estandardizada para definir um esquema de pesos. Os pesos atribuídos a cada categoria das variáveis CONF e LOC podem ser observados na Tabela 5 e na Tabela 6.

A contribuição máxima para a exactidão estimada é atribuída às observações que pertencem à categoria A5 no caso da variável CONF e A6 para a variável LOC, pois nestas situações há acordo entre a classe do mapa e a principal classe de referência (L1), e estamos perante uma zona em que não houve dúvida na interpretação da classe de referência (ICR=3) e/ou com ocupação do solo homogénea (LCR=3).

	Categorias CONF	L 1	L 2	ICR	$W_{d,CONF}$	$W_{d,MAX}$	$W_{d,RIGHT}$
<b>Acertos</b> (A)	A5	√		3	1	1	1
	A4	√		2	0.8	1	1
	A3		√	2	0.5	0	1
	A2	√		1	0.3	1	1
	A1		√	1	0.3	0	1
<b>Erros</b> (E)	E1			1	0.075	0	0
	E2			2	0	0	0
	E3			3	0	0	0

Tabela 5. Categorias CONF e respectivas contribuições para o cálculo da exactidão de  $P_{CONF}$ . O acordo entre ambos os conjuntos de dados é especificado por (√).  $W_{d,CONF}$ ,  $W_{d,MAX}$  e  $W_{d,RIGHT}$  expressam as contribuições (parciais ou totais) para o cálculo da exactidão de  $P_{CONF}$ , MAX e RIGHT respectivamente.

Pelo contrário, a contribuição para a exactidão de observações que pertencem à categoria *E3* de CONF e LOC é nula, pois nestas situações a classificação do mapa está errada, apesar de estarmos perante uma zona em que não houve dúvida na atribuição da classe de referência e/ou com uma ocupação do solo homogénea. As categorias das variáveis CONF e LOC entre estas duas situações extremas foram pesadas de acordo com a proporção de área que as classes de ocupação do solo principal e alternativa ocupam em cada uma das observações.

	Categorias LOC	L 1	L 2	LCR	$W_{d,LOC}$	$W_{d,MAX}$	$W_{d,RIGHT}$
<b>Acertos</b> (A)	A6	√		3	1	1	1
	A5	√		2	0.8	1	1
	A4		√	3	0.5	0	1
	A3		√	2	0.5	0	1
	A2	√		1	0.3	1	1
	A1		√	1	0.3	0	1
<b>Erros</b> (E)	E1			1	0.075	0	0
	E2			2	0	0	0
	E3			3	0	0	0

Tabela 6. Categorias LOC e respectivas contribuições para o cálculo da exactidão de  $P_{LOC}$ . O acordo entre ambos os conjuntos de dados é especificado por (√).  $W_{d,LOC}$ ,  $W_{d,MAX}$  e  $W_{d,RIGHT}$  expressam as contribuições (parciais ou totais) para o cálculo da exactidão de  $P_{LOC}$ , MAX e RIGHT respectivamente.

Por exemplo, nas categorias *A4* de CONF e *A5* de LOC, a classe de ocupação do solo principal (L1) deverá ocupar uma percentagem máxima de 80% (daí o peso 0.8), e nas categorias *A3* de CONF, e *A4* e *A3* de LOC, a classe de ocupação do solo

alternativa deverá ocupar uma percentagem máxima de 50% (daí os pesos 0.5). Neste exemplo, as contribuições parciais para a exactidão final do mapa, estão relacionadas com as características quantitativas da informação de referência (i.e. a proporção de área que cada classe de ocupação do solo ocupa em cada uma das observações).

No exemplo apresentado, os erros aceitáveis (*EI*) contribuem positivamente para  $P_{CONF}$  (Tabela 5) e  $P_{LOC}$  (Tabela 6). A contribuição parcial para a exactidão da categoria *EI* foi pensada da seguinte forma. A confiança que se pode ter nestas observações é reduzida, pois estas encontram-se em áreas em que a paisagem é heterogénea (pelo menos com a presença de três classes de ocupação do solo), para além de que a verdadeira classe de ocupação do solo não é nenhuma das três classes de referência que foram identificadas pelo foto-intérprete. Assumindo que a probabilidade das restantes 13 classes de ocupação do solo poder ser a verdadeira classe de ocupação do solo, existe a probabilidade de 1/13 (0.077) de que a classe de referência coincida com a classe do mapa, tendo sido utilizado o peso de 0.075.

No entanto é de salientar que a escolha dos pesos para cada uma das categorias de CONF e LOC pode ser definida pelos utilizadores. Neste sentido poderá haver utilizadores que estejam interessados por exemplo em que a confiança na classificação do mapa seja pelo menos aceitável. De forma a exemplificar esta situação efectuou-se uma análise de sensibilidade, na qual se simulou uma situação hipotética com um esquema de pesos diferente, tanto para o operador CONF como para o operador LOC. Isto irá permitir verificar a variação das medidas de exactidão  $P_{CONF}$  e  $P_{LOC}$ , com a alteração do esquema de pesos. Na Tabela 7 e na Tabela 8, são apresentadas as contribuições definidas por um utilizador, (que pretende que a confiança na classificação do mapa seja pelo menos aceitável) de cada categoria de CONF e LOC respectivamente.

Categorias CONF		L 1	L 2	ICR	$W_{d,CONF}$
<b>Acertos</b> (A)	Muito elevada confiança na classificação (A5)	√		3	1
	Elevada confiança na classificação (A4)	√		2	1
	Aceitável confiança na classificação (A3)		√	2	0.8
	Reduzida confiança na classificação (A2)	√		1	0
	Muito reduzida confiança na classificação (A1)		√	1	0
<b>Erros</b> (E)	Erro aceitável (E1)			1	0
	Erro elevado (E2)			2	0
	Erro muito elevado (E3)			3	0

Tabela 7. Categorias CONF e respectivas contribuições para o cálculo de  $P_{CONF}$  definidas por um utilizador interessado em que a confiança na classificação do mapa seja pelo menos aceitável.  $W_{d,CONF}$  expressa as contribuições (parciais ou totais) para o cálculo da exactidão de  $P_{CONF}$ .

Categorias LOC		L 1	L 2	LCR	$W_{d,LOC}$
<b>Acertos</b> (A)	Muito elevada confiança na classificação (A6)	√		3	1
	Elevada confiança na classificação (A5)	√		2	1
	Aceitável confiança na classificação (A4)		√	3	0.8
	Razoável confiança na classificação (A3)		√	2	0
	Reduzida confiança na classificação (A2)	√		1	0
	Muito reduzida confiança na classificação (A1)		√	1	0
<b>Erros</b> (E)	Erro aceitável (E1)			1	0
	Erro elevado (E2)			2	0
	Erro muito elevado (E3)			3	0

Tabela 8. Categorias LOC e respectivas contribuições para o cálculo de  $P_{LOC}$  definidas por um utilizador interessado em que a confiança na classificação do mapa seja pelo menos aceitável.  $W_{d,LOC}$  expressa as contribuições (parciais ou totais) para o cálculo da exactidão de  $P_{LOC}$ .

Outra abordagem possível é a de pode haver utilizadores que por exemplo, estejam a desenvolver estudos de vegetação e estejam mais interessados em verificar se as classes de vegetação natural apresentam uma confiança na classificação no mapa mais elevada, podendo atribuir pesos mais elevados a estas classes de ocupação do solo. Para isso o utilizador final tem apenas de substituir o valor  $W_d$  pelo peso que considera ser do seu interesse, na equação 4 (para o operador CONF) e na equação 5 (para o operador LOC).

### 3. CASO DE ESTUDO

Neste capítulo faz-se uma descrição do caso de estudo utilizado para testar a metodologia proposta. Inicialmente faz-se uma descrição da nomenclatura utilizada neste estudo, do mapa de ocupação do solo e da base de dados de referência utilizados para testar a metodologia proposta. Finalmente apresentam-se os resultados obtidos.

#### 3.1 Nomenclatura

Para testar a metodologia proposta, recorreu-se a uma nomenclatura composta por 16 classes de ocupação do solo (Tabela 9), baseada na nomenclatura LANDEO desenvolvida por Carrão *et al.* (2007). A nomenclatura LANDEO é simples, clara, com limites de classes bem definidos, e com a possibilidade de ser aplicada à produção de mapas de ocupação do solo a diferentes escalas de análise.

Código	Classe de ocupação do solo	Código	Classe de ocupação do solo
5	Áreas húmidas	242	Vinhas
6	Solo nu / vegetação esparsa	34	Vegetação arbustiva
7	Água	35	Vegetação herbácea
11	Áreas artificiais contínuas	38	Áreas ardidas e cortes rasos
12	Áreas artificiais descontínuas	311	Floresta de folhosas
21	Culturas de sequeiro	312	Áreas agro-florestais
22	Culturas de regadio	321	Floresta de resinosas
23	Arrozais	331	Floresta mista

Tabela 9. Nomenclatura de ocupação do solo LANDEO (Fonte: Carrão *et al.* 2009).

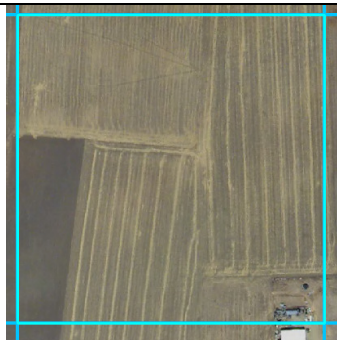
O objectivo do desenvolvimento desta nomenclatura foi o de esta permitir uma comparação e harmonização de cartografias produzidas com diferentes nomenclaturas (e.g. IGBP (*International Geosphere-Biosphere Programme*), CLC (*CORINE Land Cover*) e GLC (*Global Land Cover*)). Na Figura 4, pode ser consultado um exemplo de cada uma das classes de ocupação do solo da nomenclatura utilizada neste estudo.



Classe 11



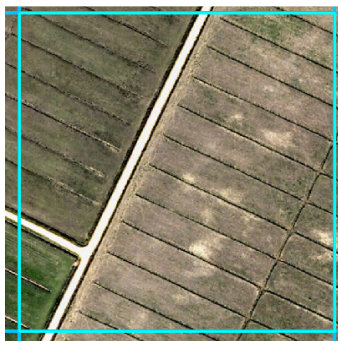
Classe 12



Classe 21



Classe 22



Classe 23



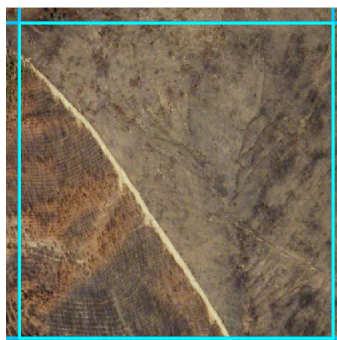
Classe 242



Classe 34



Classe 35



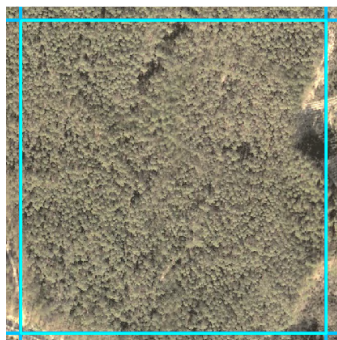
Classe 38



Classe 311



Classe 312



Classe 321

(continuação na página seguinte)



(continuação da página anterior)

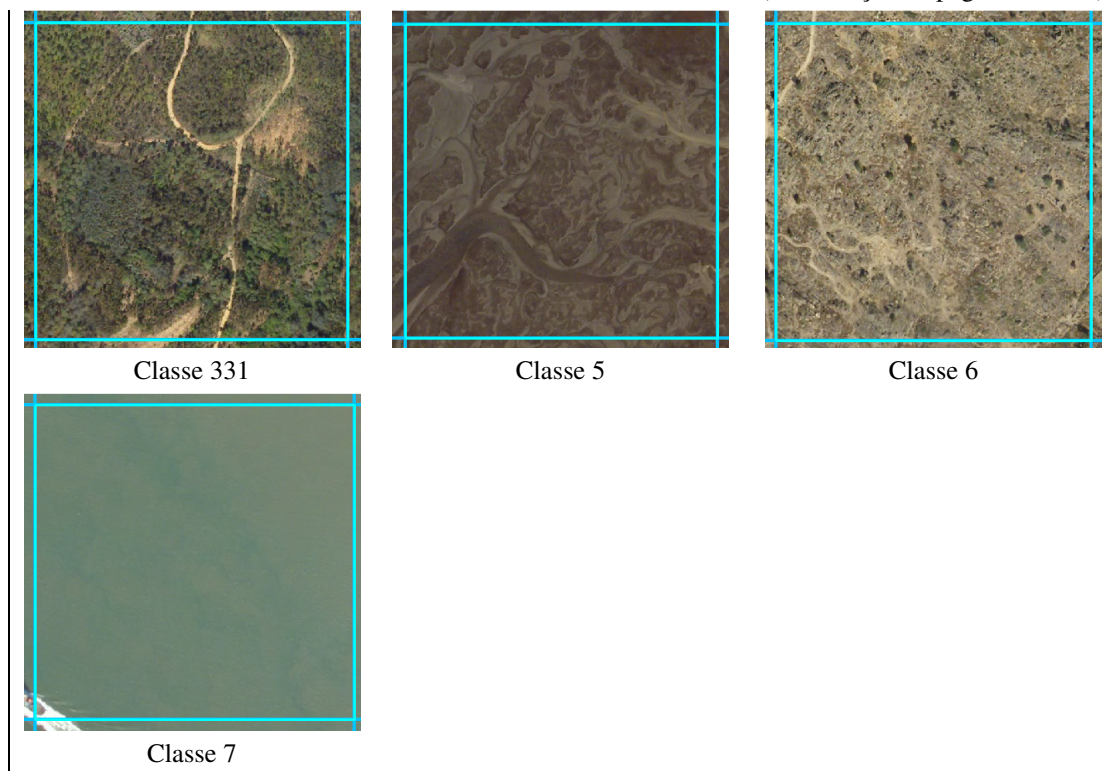


Figura 4. Exemplo de cada uma das classes de ocupação do solo da nomenclatura utilizada no estudo proposto.

### 3.2 Mapa de ocupação do solo para Portugal Continental

O mapa de ocupação do solo utilizado foi derivado da classificação automática de imagens MERIS Full Resolution. Estas imagens possuem uma resolução espacial de 300 m e 13 bandas espectrais. O mapa foi produzido através da combinação de dois métodos de classificação distintos, aplicados a um conjunto de seis compósitos bimensais do ano de 2005. Um Classificador Linear Discriminante foi utilizado na classificação das classes sem dinâmica intra-anual (e.g. áreas artificiais). Por outro lado, classes como a de áreas áridas e cortes rasos com elevada dinâmica intra-anual, foram classificadas com uma técnica de diferenciação do índice de vegetação (Carrão *et al.* 2009). A escolha de mapas produzidos a partir de imagens derivadas destes sensores está relacionada com o facto de existirem produtos globais de ocupação do solo derivados dos mesmos, sendo também estes produtos, alvo de estudo tanto pelo GOF-C-GOLD e CEOS-WGCV.

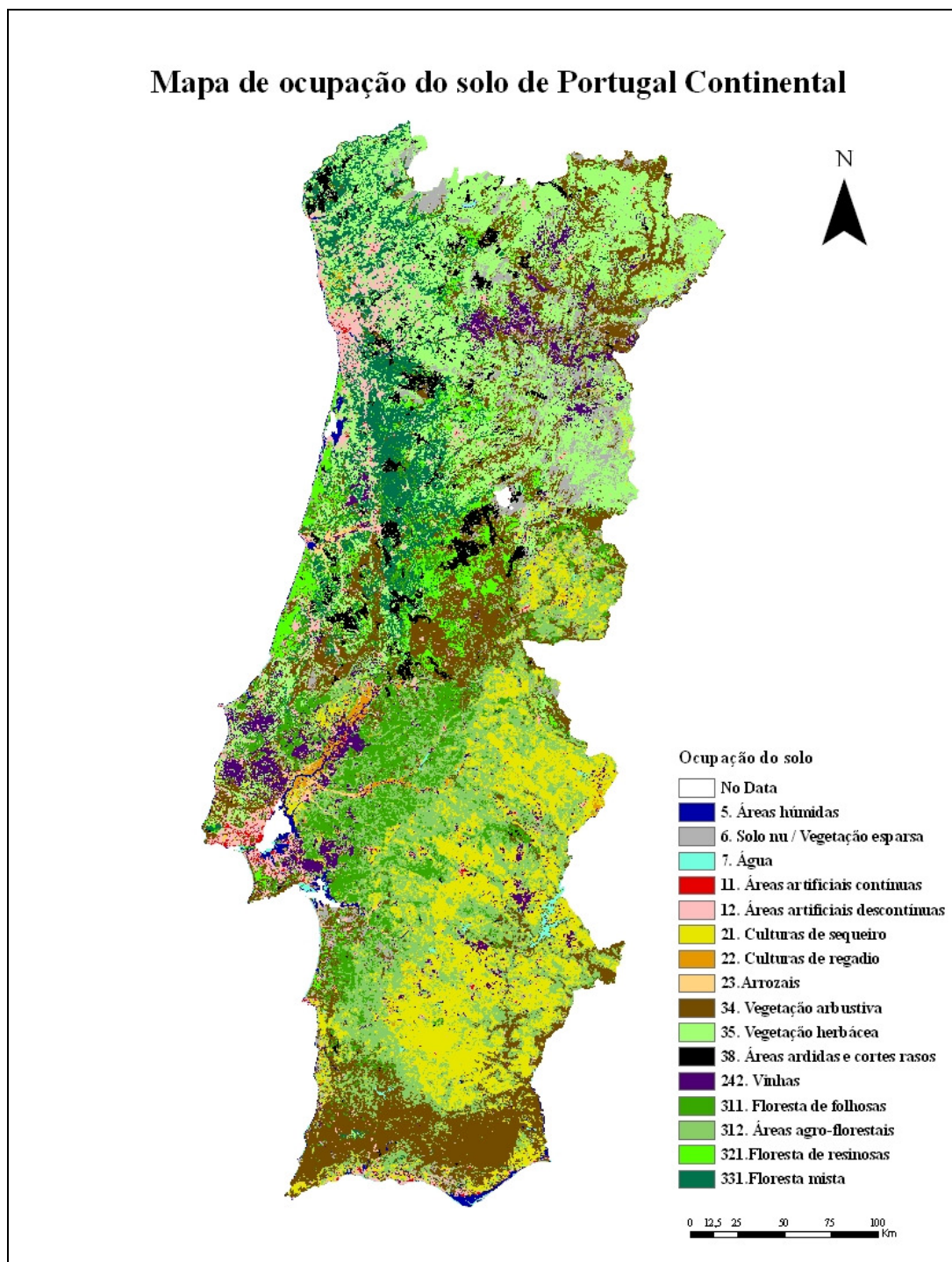


Figura 5. Mapa de ocupação do solo produzido a partir do compósito bimensal derivado do sensor MERIS (Adaptado de Carrão *et al.* 2009).

A título de exemplo no GDR do IGP, foram já desenvolvidos alguns estudos no desenvolvimento de metodologias para produção de mapas de ocupação do solo a partir das mesmas imagens utilizadas para mapas globais, com o intuito de verificar a sua possível aplicação em programas nacionais de produção de mapas de ocupação

do solo (Carrão *et al.* 2009). A motivação para a realização destes estudos está relacionada com a iniciativa GlobCover (Arino *et al.* 2007). Pretende-se com a iniciativa GlobCover, desenvolver um serviço que permita a produção de um mapa de ocupação do solo à escala global, relativo aos anos de 2005 e 2006, com uma resolução espacial de 300 m. É esperado que este novo produto complemente os produtos existentes (e.g. GLC2000 com uma resolução espacial de 1 Km), devido a possuir uma resolução espacial mais elevada. Na Figura 5, é apresentado o mapa de ocupação do solo de Portugal Continental, utilizado para testar a metodologia proposta de avaliação da exactidão temática.

### **3.3 Base de dados de referência**

Para elaborar a base de dados de referência recorreu-se a uma amostragem aleatória estratificada. Os estratos foram baseados nas classes de ocupação do solo definidas numa versão preliminar do mapa utilizado neste estudo, tendo sido recolhidas posteriormente um conjunto de unidades amostrais para cada classe. A utilização de amostragem aleatória estratificada deveu-se ao facto de se pretender que todas as classes de ocupação do solo tivessem a mesma probabilidade de serem amostradas. Na Figura 6 é apresentada a distribuição espacial da amostra pelo território de Portugal Continental.

A informação de referência apresentada na Tabela 2 foi recolhida por quatro foto-intérpretes, através da foto-interpretação (em cada um dos blocos de 3x3 pixéis em torno do pixel central) de um conjunto de imagens aéreas de elevada resolução espacial referentes aos anos de 2004, 2005 e 2006 adquiridas pelo IGP e a Direcção Geral dos Recursos Florestais (DGRF), que cobrem a totalidade do território de Portugal Continental. Estas imagens aéreas possuem uma resolução espacial de 0.5 m e quatro bandas espectrais (três bandas na região do visível e uma na região do infra-vermelho próximo).

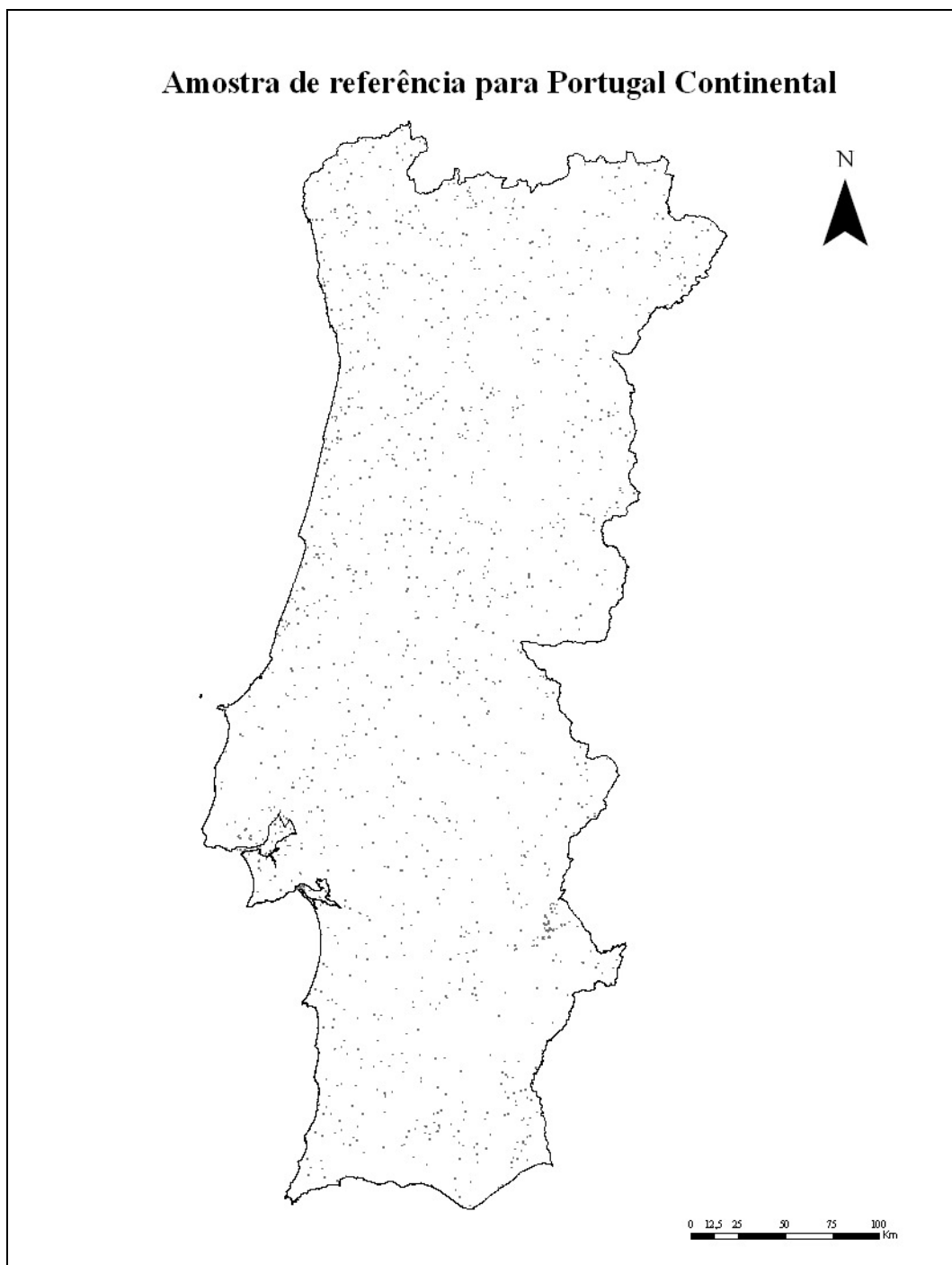


Figura 6. Distribuição espacial da amostra de referência pelo território de Portugal Continental.

Na Tabela 10 é apresentado o número de observações recolhidas para cada classe de ocupação do solo como ocupação do solo principal (L1) e secundária (L2).

<b>Código</b>	<b>Classe de ocupação do solo</b>	<b>L1</b>	<b>L2</b>
<b>5</b>	Áreas húmidas	21	45
<b>6</b>	Solo nu / vegetação esparsa	63	48
<b>7</b>	Água	126	120
<b>11</b>	Áreas artificiais contínuas	38	30
<b>12</b>	Áreas artificiais descontínuas	84	102
<b>21</b>	Culturas de sequeiro	178	143
<b>22</b>	Culturas de regadio	65	73
<b>23</b>	Arrozais	40	28
<b>34</b>	Vegetação arbustiva	324	343
<b>35</b>	Vegetação herbácea	346	232
<b>38</b>	Áreas ardidas e cortes rasos	62	36
<b>242</b>	Vinhas	42	33
<b>311</b>	Floresta de folhosas	89	115
<b>312</b>	Áreas agro-florestais	142	252
<b>321</b>	Floresta de resinosas	133	162
<b>331</b>	Floresta mista	99	86

Tabela 10. Número de observações recolhidas para cada classe de ocupação do solo, como ocupação do solo principal (L1) e secundária (L2).

### 3.4 Resultados

Nesta secção são apresentados os resultados obtidos com a metodologia proposta de avaliação da exactidão temática. O ICR médio na atribuição das classes de ocupação do solo de referência e o ICR médio para acertos e erros entre mapa e referência é mostrado na Tabela 11. O ICR médio varia bastante de classe para classe, indicando as classes de ocupação do solo às quais foi atribuída elevada confiança na referência (e.g. áreas húmidas (5), água (7) e arrozais (23)) e às quais foi atribuída pouca confiança (e.g. vegetação esparsa / solo nu (6) e vegetação herbácea (5)). A elevada incerteza associada às classes de vegetação esparsa / solo nu (6) e vegetação herbácea (5) é devida à difícil observação do limiar de percentagem de coberto que foi definido para distinguir ambas as classes. No geral o ICR médio é superior quando existe acordo entre mapa e referência (Tabela 11), reflectindo assim a relação esperada entre a concordância entre ambos os conjuntos de dados e a confiança na atribuição das classes de ocupação do solo. A confiança na atribuição de uma classe de ocupação do solo é reduzida, quando existem várias classes de ocupação do solo numa determinada observação da amostra, sendo de esperar que os erros do mapa

aconteçam na sua maioria nestas situações. A única excepção é relativa à classe de vegetação herbácea (35), em que o ICR médio é mais elevado para os erros do que para os acertos. Este aspecto pode ser explicado devido ao ICR médio para esta classe ser muito reduzido (1.27).

Classes de ocupação do solo	Mapa <i>versus</i> Referência		
	Dados de referência	Erros	Acertos
	ICR médio	ICR médio	ICR médio
Áreas húmidas	2.33	2.00	2.37
Solo nu / vegetação esparsa	1.68	1.43	1.84
Água	2.65	-	2.94
Áreas artificiais contínuas	2.29	1.33	2.48
Áreas artificiais descontínuas	2.02	1.90	2.16
Culturas de sequeiro	1.89	1.54	2.34
Culturas de regadio	1.86	2.00	2.71
Arrozais	2.55	2.33	2.67
Vegetação arbustiva	1.74	1.60	1.91
Vegetação herbácea	1.27	1.55	1.26
Áreas ardidas e cortes rasos	2.11	1.55	2.53
Vinhas	1.74	1.89	2.04
Floresta de folhosas	1.99	1.42	2.37
Áreas agro-florestais	1.86	1.57	2.29
Floresta de resinosas	2.20	2.04	2.65
Floresta mista	1.86	1.68	1.98
<b>Total</b>	<b>1.85</b>	<b>1.62</b>	<b>2.13</b>

Tabela 11. Índices de confiança de interpretação médios (ICR) para os dados de referência, e acertos e erros entre mapa e referência.

Na Tabela 12, são apresentados os valores médios de confiança na localização (LCR) da base de dados de referência, bem como dos acertos e erros entre mapa e referência. Tal como para o índice de confiança de interpretação, os valores do índice de confiança de localização também não são muito elevados, quer globalmente, como ao nível da classe. Este aspecto revela que no geral, a maioria das observações da amostra de referência, está inserida em áreas em que a ocupação do solo é heterogénea. Deste modo, fica assim evidente a fragmentação da paisagem de Portugal Continental, quando representada com uma unidade mínima cartográfica de 9 ha (área ocupada por um *pixel* MERIS). Esta fragmentação é assim o factor responsável pela dificuldade na interpretação da ‘verdadeira’ ocupação do solo de

referência. No entanto, para as classes de áreas húmidas (5) e culturas de regadio (22), existe um valor de confiança de localização médio mais elevado para os erros quando comparado com os acertos entre ambos os conjuntos de dados.

Classes de ocupação do solo	Dados de referência	Mapa versus Referência	
		Erros	Acertos
	LCR médio	LCR médio	
Áreas húmidas	2.10	2.27	2.08
Solo nu / vegetação esparsa	2.14	1.68	2.18
Água	2.34	-	2.40
Áreas artificiais contínuas	2.18	1.67	2.30
Áreas artificiais descontínuas	1.75	1.76	1.87
Culturas de sequeiro	1.95	2.07	2.23
Culturas de regadio	1.92	2.50	2.15
Arrozais	2.35	2.00	2.54
Vegetação arbustiva	1.80	1.90	1.87
Vegetação herbácea	1.68	1.58	1.52
Áreas ardidas e cortes rasos	2.08	1.76	2.35
Vinhas	1.50	1.79	1.61
Floresta de folhosas	2.07	2.04	2.29
Áreas agro-florestais	2.23	1.89	2.21
Floresta de resinosas	2.18	1.61	2.31
Floresta mista	1.77	1.71	1.68
<b>Total</b>	1.94	1.81	1.99

Tabela 12. Índices de confiança de localização médios (LCR) para os dados de referência, e acertos e erros entre mapa e referência.

Para as classes de vegetação arbustiva (34), vegetação herbácea (35), vinhas (242) e floresta mista (331), verificou-se a mesma tendência, mas com uma diferença entre erros e acertos mais reduzida. Para estas classes, pode verificar-se que o LCR médio é muito reduzido, ou seja, estas classes estão na sua grande maioria em zonas em que a paisagem é heterogénea.

Na Tabela 13 apresentam-se os índices de exactidão temática MAX, RIGHT e  $P_{CONF}$  para o mapa em estudo. A proporção de área que é ocupada por cada uma das categorias da variável CONF é apresentada também na Tabela 13.

<b>Classes de ocupação do solo</b>	<b>Observações</b>	<b><i>E3</i></b>	<b><i>E2</i></b>	<b><i>E1</i></b>	<b><i>A1</i></b>	<b><i>A2</i></b>	<b><i>A3</i></b>	<b><i>A4</i></b>	<b><i>A5</i></b>	<b>MAX</b>	<b>RIGHT</b>	<b><math>P_{CONF}</math></b>
<b>5</b>	49	0.082	0.061	0.082	0.102	0.082	0.265	0.082	0.245	0.39	0.78	0.50
<b>6</b>	92	0.054	0.076	0.272	0.141	0.152	0.043	0.174	0.087	0.41	0.60	0.36
<b>7</b>	90	0.000	0.000	0.000	0.044	0.022	0.011	0.011	0.911	0.94	1.00	0.95
<b>11</b>	26	0.000	0.038	0.077	0.038	0.154	0.038	0.115	0.538	0.81	0.88	0.71
<b>12</b>	105	0.067	0.048	0.086	0.124	0.171	0.076	0.162	0.267	0.60	0.80	0.53
<b>21</b>	171	0.035	0.099	0.181	0.076	0.140	0.029	0.099	0.339	0.58	0.68	0.51
<b>22</b>	22	0.045	0.000	0.045	0.045	0.045	0.091	0.136	0.591	0.77	0.91	0.78
<b>23</b>	27	0.074	0.000	0.037	0.000	0.074	0.000	0.148	0.667	0.89	0.89	0.81
<b>34</b>	335	0.039	0.069	0.134	0.116	0.251	0.054	0.137	0.200	0.59	0.76	0.46
<b>35</b>	318	0.044	0.082	0.186	0.142	0.365	0.075	0.085	0.022	0.47	0.69	0.29
<b>38</b>	67	0.075	0.119	0.299	0.015	0.075	0.015	0.075	0.328	0.48	0.51	0.44
<b>242</b>	64	0.125	0.141	0.172	0.141	0.141	0.000	0.125	0.156	0.42	0.56	0.35
<b>311</b>	81	0.049	0.037	0.235	0.099	0.148	0.074	0.025	0.333	0.51	0.68	0.48
<b>312</b>	172	0.035	0.105	0.169	0.209	0.081	0.087	0.116	0.198	0.40	0.69	0.43
<b>321</b>	114	0.079	0.053	0.070	0.123	0.070	0.053	0.079	0.474	0.62	0.80	0.63
<b>331</b>	119	0.067	0.084	0.168	0.101	0.210	0.034	0.134	0.202	0.55	0.68	0.43
<b>Exactidão global</b>										0.51	0.70	0.44

Tabela 13. Proporção de observações em cada categoria CONF por classe de ocupação do solo, e as medidas de exactidão temática MAX, RIGHT e  $P_{CONF}$ . As medidas de exactidão apresentadas, são análogas à exactidão do utilizador, porque o divisor é baseado no número de observações presentes no mapa para cada classe de ocupação do solo. Os acrónimos das categorias CONF são: *E3*=erro muito elevado; *E2*=erro elevado; *E1*=erro aceitável; *A1*=muito reduzida confiança na classificação; *A2*=reduzida confiança na classificação; *A3*=aceitável confiança na classificação; *A4*=elevada confiança na classificação; *A5*=muito elevada confiança na classificação.



Observando a Tabela 13, pode verificar-se que RIGHT apresenta os valores de exactidão mais elevados, tanto no global como por classe, comparativamente a MAX e  $P_{CONF}$ . De facto RIGHT apresenta uma contribuição máxima em todas as categorias de acertos de CONF (Tabela 5), sendo assim RIGHT uma medida de exactidão que dá uma perspectiva demasiado optimística da exactidão do mapa em estudo.

Relativamente à exactidão global de  $P_{CONF}$ , verifica-se que é inferior à de MAX (0.44 e 0.51, respectivamente). De modo semelhante, a exactidão por classe de  $P_{CONF}$  é também inferior a MAX. De facto o esquema de pesos escolhido é bastante penalizador, para as categorias CONF que têm valores de ICR reduzidos (i.e. A4 e A2 representam acertos entre mapa e referência, mas as suas contribuições para a exactidão são 0.8 e 0.3 respectivamente). Neste sentido é assim esperado que o normal seja que os valores de exactidão de MAX sejam superiores a  $P_{CONF,m}$ , nos casos em que haja uma elevada proporção de acertos, nas categorias A4 e A2 da variável CONF. Na Tabela 5 são apresentadas as contribuições parciais de A4 e A2 de  $P_{CONF,m}$ , mais reduzidas do que as contribuições parciais de MAX.

No geral, para o caso de estudo apresentado, a confiança na atribuição das classes de ocupação do solo tende a ser reduzida. Este aspecto é evidenciado pela relativa reduzida proporção de observações que pertencem às categorias A5 e A4 de CONF, como também pelo valor máximo que  $P_{CONF,m}$  pode atingir (Tabela 14). Apenas cinco classes de ocupação do solo apresentam um valor máximo possível para  $P_{CONF,m}$ , superior a 0.8. A classe de áreas húmidas (5) é a única em que  $P_{CONF}$  excede largamente MAX. Para esta classe de ocupação do solo, existe uma elevada contribuição para a exactidão  $P_{CONF}$ , da categoria A3 de CONF (Tabela 14) (A3, representam acordo entre a classe de ocupação do solo alternativa e a classe do mapa). Nesta categoria, MAX apresenta um peso de 0 (Tabela 5).

Classes de ocupação do solo	<i>EI</i> (0.075)	<i>A1</i> (0.3)	<i>A2</i> (0.3)	<i>A3</i> (0.5)	<i>A4</i> (0.8)	<i>A5</i> (1)	$P_{CONF}$	Acordo	$P_{CONF}$ MAX
5	0.006	0.031	0.024	0.133	0.065	0.245	0.50	0.46	0.73
6	0.020	0.042	0.046	0.022	0.139	0.087	0.36	0.36	0.69
7	0.000	0.013	0.007	0.006	0.009	0.911	0.95	0.97	0.97
11	0.006	0.012	0.046	0.019	0.092	0.538	0.71	0.73	0.86
12	0.006	0.037	0.051	0.038	0.130	0.267	0.53	0.54	0.78
21	0.014	0.023	0.042	0.015	0.080	0.339	0.51	0.52	0.79
22	0.003	0.014	0.014	0.045	0.109	0.591	0.78	0.76	0.87
23	0.003	0.000	0.022	0.000	0.119	0.667	0.81	0.82	0.94
34	0.010	0.035	0.075	0.027	0.110	0.200	0.46	0.49	0.76
35	0.014	0.042	0.109	0.038	0.068	0.022	0.29	0.34	0.69
38	0.022	0.004	0.022	0.007	0.060	0.328	0.44	0.43	0.78
242	0.013	0.042	0.042	0.000	0.100	0.156	0.35	0.37	0.76
311	0.018	0.030	0.044	0.037	0.020	0.333	0.48	0.48	0.75
312	0.013	0.063	0.024	0.044	0.093	0.198	0.43	0.42	0.71
321	0.005	0.037	0.021	0.026	0.063	0.474	0.63	0.63	0.84
331	0.013	0.030	0.063	0.017	0.108	0.202	0.43	0.46	0.76
<b>Exactidão global</b>							0.44	0.45	0.74

Tabela 14. Contribuições ponderadas de cada categoria de CONF para o valor de exactidão  $P_{CONF}$ . O valor do peso para cada categoria de CONF está listado entre parênteses por baixo de cada categoria. A coluna 'Acordo', indica a contribuição para  $P_{CONF}$ , de todas as observações que correspondem a acertos entre a classe de referência e mapa (i.e. categorias *A1*, *A2*, *A3*, *A4* e *A5*).  $P_{CONF}$  MAX, representa o valor máximo possível de  $P_{CONF}$ , se todas as observações em cada categoria de CONF, representarem acordo entre a classe do mapa e a classe de referência principal (L1). Os acrónimos das categorias CONF são: *E3*=erro muito elevado; *E2*=erro elevado; *E1*=erro aceitável; *A1*=muito reduzida confiança na classificação; *A2*=reduzida confiança na classificação; *A3*=aceitável confiança na classificação; *A4*=elevada confiança na classificação; *A5*=muito elevada confiança na classificação.

Na Tabela 14 são apresentadas as contribuições parciais para o valor de exactidão  $P_{CONF}$ , de cada uma das categorias de CONF. Esta tabela permite ao utilizador final perceber se para os valores de exactidão, contribuem observações nas quais foi atribuída muita confiança ou pouca confiança à classe de ocupação do solo principal. O formato da Tabela 14 permite também uma reinterpretação dos valores de exactidão, bastando fazer uma alteração do esquema de pesos segundo os interesses do utilizador. Para a alteração do esquema de pesos presente na Tabela 5, as contribuições para  $P_{CONF}$  de cada categoria CONF podem ser calculadas, inserindo o esquema de pesos definido pelo utilizador na equação (6).

Na Tabela 15 são apresentadas as proporções observadas para cada uma das classes de ocupação do solo em cada uma das categorias LOC, assim como os valores de exactidão temática obtidos com os operadores MAX, RIGHT e  $P_{LOC}$ . RIGHT apresenta valores de exactidão quer no global quer por classe superiores a MAX e  $P_{LOC}$ , à semelhança do que sucede com  $P_{CONF}$ . De facto RIGHT contempla uma contribuição máxima em todas as categorias de acerto LOC (Tabela 6). No entanto, esta medida de exactidão sobrestima a exactidão, pois considera como acertos todos os casos em que a classificação do mapa coincide com uma das duas classes de referência (L1 ou L2), e não tem em conta a influência da incerteza posicional da informação de referência.

A exactidão global de  $P_{LOC}$  é mais reduzida que a de MAX (0.44 e 0.51 respectivamente). De modo semelhante a exactidão temática por classe é por regra mais elevada em MAX do que  $P_{LOC,m}$ . De facto, a proporção de observações que pertencem às categorias A6 e A5 de LOC que contemplam áreas homogéneas e no limite entre classes (LCR igual a 3 e 2 respectivamente) é muito elevada (i.e. nas categorias A6 e A5 existe acordo entre a classificação do mapa e a classe de referência principal (L1), mas a contribuição para a exactidão de  $P_{LOC,m}$  é de 1 para a primeira e 0.8 para a segunda, enquanto que para MAX existe uma contribuição de 1 nos dois casos). As classes de ocupação do solo de áreas húmidas (5) e áreas agro-florestais (312), são as únicas que apresentam um valor de  $P_{LOC,m}$  superior a MAX. No caso da classe de áreas húmidas (5), este facto deve-se à elevada proporção de observações que pertencem à categoria A3, ou seja que correspondem a acertos entre a classificação do mapa e a ocupação do solo de referência alternativa (L2) (MAX

Classes de ocupação do solo	Observações	<i>E3</i>	<i>E2</i>	<i>E1</i>	<i>A1</i>	<i>A2</i>	<i>A3</i>	<i>A4</i>	<i>A5</i>	<i>A6</i>	MAX	RIGHT	$P_{LOC}$
5	49	0.061	0.163	0.000	0.041	0.061	0.286	0.061	0.224	0.102	0.39	0.78	0.49
6	92	0.065	0.141	0.196	0.043	0.109	0.043	0.098	0.141	0.163	0.41	0.60	0.41
7	90	0.000	0.000	0.000	0.000	0.022	0.056	0.000	0.500	0.422	0.94	1.00	0.86
11	26	0.038	0.000	0.077	0.038	0.154	0.038	0.000	0.192	0.462	0.81	0.88	0.70
12	105	0.038	0.076	0.086	0.076	0.229	0.095	0.029	0.200	0.171	0.60	0.80	0.49
21	171	0.117	0.105	0.094	0.035	0.123	0.053	0.018	0.158	0.298	0.58	0.68	0.51
22	22	0.045	0.045	0.000	0.045	0.136	0.091	0.000	0.318	0.318	0.77	0.91	0.67
23	27	0.037	0.037	0.037	0.000	0.000	0.000	0.000	0.407	0.481	0.89	0.89	0.81
34	335	0.054	0.110	0.078	0.069	0.257	0.045	0.057	0.164	0.167	0.59	0.76	0.45
35	318	0.028	0.123	0.160	0.104	0.321	0.072	0.041	0.097	0.053	0.47	0.69	0.33
38	67	0.075	0.224	0.194	0.015	0.090	0.000	0.015	0.119	0.269	0.48	0.51	0.42
242	64	0.063	0.219	0.156	0.078	0.234	0.016	0.047	0.141	0.047	0.42	0.56	0.30
311	81	0.111	0.111	0.099	0.012	0.111	0.062	0.099	0.173	0.222	0.51	0.68	0.49
312	172	0.076	0.122	0.110	0.110	0.058	0.070	0.116	0.140	0.198	0.40	0.69	0.46
321	114	0.026	0.070	0.105	0.079	0.088	0.053	0.044	0.167	0.368	0.62	0.80	0.61
331	119	0.034	0.160	0.126	0.076	0.244	0.050	0.008	0.210	0.092	0.55	0.68	0.40
<b>Exactidão global</b>											0.51	0.70	0.44

Tabela 15. Proporção de observações em cada categoria LOC por classe de ocupação do solo, e as medidas de exactidão temática MAX, RIGHT e  $P_{LOC}$ . As medidas de exactidão apresentadas, são análogas à exactidão do utilizador, porque o divisor é baseado no número de observações presentes no mapa para cada classe de ocupação do solo. Os acrónimos das categorias LOC são: *E3*=erro muito elevado; *E2*=erro elevado; *E1*=erro aceitável; *A1*=muito reduzida confiança na classificação; *A2*=reduzida confiança na classificação; *A3*=razoável confiança na classificação; *A4*=aceitável confiança na classificação; *A5*=elevada confiança na classificação; *A6*=elevada confiança na classificação.

apresenta um peso de 0 nesta categoria (Tabela 6)). Por outro lado a soma das proporções em *A4*, *A3* e *A1* para as duas classes mencionadas é bastante elevada (0.388 para a classe de áreas húmidas e 0.378 para a classe de áreas agro-florestais). Note-se que MAX tem peso 0 nestas três categorias LOC. A influência do esquema de pesos nas medidas de exactidão temáticas é bastante evidente relativamente à classe de vegetação herbácea (35) e vegetação arbustiva (34). Note-se que a elevada proporção de observações na categoria LOC *A2*, faz com que MAX (peso 1 em *A2*), seja bastante superior relativamente a  $P_{LOC,m}$  (peso 0.3 em *A2*), para as classes mencionadas anteriormente.

Comparando os valores de  $P_{CONF}$  (Tabela 14) e  $P_{LOC}$  (Tabela 16), pode verificar-se que  $P_{CONF}$  apresenta geralmente sempre valores superiores a  $P_{LOC}$ , mas muito semelhantes. De facto esta semelhança indica que a incerteza humana aquando da foto-interpretação das classes de ocupação do solo de referência está intimamente relacionada com a heterogeneidade da paisagem. As classes de ocupação do solo com maior diferença entre  $P_{CONF}$  e  $P_{LOC}$  são as classes de água (7), áreas urbanas descontínuas (12) e culturas de regadio (22). Este facto deve-se essencialmente ao número de observações que pertencem à categoria *A5* (muito elevada confiança na classificação) de CONF ser bastante superior quando comparada com a categoria *A6* (muito elevada confiança na classificação) de LOC.

Classes de ocupação do solo	<i>E1</i> (0.075)	<i>A1</i> (0.3)	<i>A2</i> (0.3)	<i>A3</i> (0.5)	<i>A4</i> (0.5)	<i>A5</i> (0.8)	<i>A6</i> (1)	$P_{LOC}$	Acordo
5	0.000	0.012	0.018	0.143	0.031	0.180	0.102	0.49	0.49
6	0.015	0.013	0.033	0.022	0.049	0.113	0.163	0.41	0.39
7	0.000	0.000	0.007	0.028	0.000	0.400	0.422	0.86	0.86
11	0.006	0.012	0.046	0.019	0.000	0.154	0.462	0.70	0.69
12	0.006	0.023	0.069	0.048	0.014	0.160	0.171	0.49	0.48
21	0.007	0.011	0.037	0.026	0.009	0.126	0.298	0.51	0.51
22	0.000	0.014	0.041	0.045	0.000	0.255	0.318	0.67	0.67
23	0.003	0.000	0.000	0.000	0.000	0.326	0.481	0.81	0.81
34	0.006	0.021	0.077	0.022	0.028	0.131	0.167	0.45	0.45
35	0.012	0.031	0.096	0.036	0.020	0.078	0.053	0.33	0.32
38	0.015	0.004	0.027	0.000	0.007	0.096	0.269	0.42	0.40
242	0.012	0.023	0.070	0.008	0.023	0.113	0.047	0.30	0.28
311	0.007	0.004	0.033	0.031	0.049	0.138	0.222	0.49	0.48
312	0.008	0.033	0.017	0.035	0.058	0.112	0.198	0.46	0.45
321	0.008	0.024	0.026	0.026	0.022	0.133	0.368	0.61	0.60
331	0.009	0.023	0.073	0.025	0.004	0.168	0.092	0.40	0.39
<b>Exactidão global</b>								0.44	0.43

Tabela 16. Contribuições ponderadas de cada categoria de LOC para o valor de exactidão  $P_{LOC}$ . O valor do peso para cada categoria de LOC está listado entre parênteses por baixo de cada categoria. A coluna ‘Acordo’, indica a contribuição para  $P_{LOC}$ , de todas as observações que correspondem a acertos entre a classe de referência e mapa (i.e. categorias *A1*, *A2*, *A3*, *A4*, *A5* e *A6*). Os acrónimos das categorias LOC são: *E3*=erro muito elevado; *E2*=erro elevado; *E1*=erro aceitável; *A1*=muito reduzida confiança na classificação; *A2*=reduzida confiança na classificação; *A3*=razoável confiança na classificação; *A4*=aceitável confiança na classificação; *A5*=elevada confiança na classificação; *A6*=muito elevada confiança na classificação.

### 3.4.1 Análise de sensibilidade

Na Tabela 17, são apresentadas as contribuições parciais para a exactidão de  $P_{CONF,USER}$  de cada categoria de CONF em que a confiança na classificação do mapa é pelo menos aceitável. De modo geral a exactidão global e por classe calculada com os pesos definidos pelo utilizador ( $P_{CONF,USER}$ ) é inferior à de  $P_{CONF}$ .

Classes de ocupação do solo	A3 (0.8)	A4 (1)	A5 (1)	$P_{CONF,USER}$	$P_{CONF}$
5	0.212	0.082	0.245	0.54	0.50
6	0.035	0.174	0.087	0.30	0.36
7	0.009	0.011	0.911	0.93	0.95
11	0.031	0.115	0.538	0.68	0.71
12	0.061	0.162	0.267	0.49	0.53
21	0.023	0.099	0.339	0.46	0.51
22	0.073	0.136	0.591	0.80	0.78
23	0.000	0.148	0.667	0.81	0.81
34	0.043	0.137	0.200	0.38	0.46
35	0.060	0.085	0.022	0.17	0.29
38	0.012	0.075	0.328	0.41	0.44
242	0.000	0.125	0.156	0.28	0.35
311	0.059	0.025	0.333	0.42	0.48
312	0.070	0.116	0.198	0.38	0.43
321	0.042	0.079	0.474	0.59	0.63
331	0.027	0.134	0.202	0.36	0.43
<b>Exactidão global</b>				0.37	0.44

Tabela 17. Contribuições ponderadas de cada categoria de CONF, definidas por um utilizador para o valor de exactidão  $P_{CONF,USER}$ . O valor do peso para cada categoria de CONF está listado entre parênteses por baixo de cada categoria. Os acrónimos das categorias CONF são: A3=aceitável confiança na classificação; A4=elevada confiança na classificação; A5=muito elevada confiança na classificação.

As únicas excepções são relativas às classes de áreas húmidas (5) e de culturas de regadio (22) que apresentaram incrementos de 4% e 2% respectivamente. Este facto deve-se ao elevado número de observações que a classe de áreas húmidas (5) apresenta na categoria A3 de CONF (Tabela 13) e ao respectivo peso para esta categoria ser de 0.8 ao invés do peso 0.5 anteriormente utilizado (Tabela 5). Já no caso da classe de culturas de regadio (22), o valor mais elevado de  $P_{CONF,USER}$ , deve-se à relativa elevada proporção de observações que pertencem à categoria A4 de CONF que agora apresenta um peso 1. A classe que sofreu a maior redução de exactidão foi a de vegetação herbácea (35) (uma diferença de 12% relativamente a  $P_{CONF}$ ). De facto esquema de pesos definido pelo utilizador é bastante penalizador para as categorias CONF em que é considerado haver pouca confiança na

classificação (i.e. categorias A1 e A2 de CONF que nesta análise têm peso 0 e nas quais a classe de vegetação herbácea (35), apresenta um elevado número de observações (Tabela 13)).

Na Tabela 18, são apresentadas as contribuições parciais para a exactidão de  $P_{LOC,USER}$  de cada categoria de LOC em que a confiança na classificação do mapa é pelo menos aceitável. À semelhança de  $P_{CONF,USER}$  relativamente a  $P_{CONF}$ , a exactidão global e por classe de  $P_{LOC,USER}$  é inferior à de  $P_{LOC}$ .

Classes de ocupação do solo	A4 (0.8)	A5 (1)	A6 (1)	$P_{LOC,USER}$	$P_{LOC}$
5	0.049	0.224	0.102	0.38	0.49
6	0.078	0.141	0.163	0.38	0.41
7	0.000	0.500	0.422	0.92	0.86
11	0.000	0.192	0.462	0.65	0.70
12	0.023	0.200	0.171	0.39	0.49
21	0.014	0.158	0.298	0.47	0.51
22	0.000	0.318	0.318	0.64	0.67
23	0.000	0.407	0.481	0.89	0.81
34	0.045	0.164	0.167	0.38	0.45
35	0.033	0.097	0.053	0.18	0.33
38	0.012	0.119	0.269	0.40	0.42
242	0.038	0.141	0.047	0.23	0.30
311	0.079	0.173	0.222	0.47	0.49
312	0.093	0.140	0.198	0.43	0.46
321	0.035	0.167	0.368	0.57	0.61
331	0.007	0.210	0.092	0.31	0.40
<b>Exactidão global</b>				0.37	0.44

Tabela 18. Contribuições ponderadas de cada categoria de LOC, definidas por um utilizador para o valor de exactidão  $P_{LOC,USER}$ . O valor do peso para cada categoria de LOC está listado entre parênteses por baixo de cada categoria. Os acrónimos das categorias LOC são: A4=aceitável confiança na classificação; A5=elevada confiança na classificação; A6=muito elevada confiança na classificação.

As únicas excepções a esta regra são relativas às classes de água (7) e arrozais (23), em que há um aumento na exactidão de  $P_{LOC,USER}$  de 6% e 8% respectivamente. Apesar de ambas estas classes não apresentarem observações na categoria LOC A4, possuem um elevado número de observações nas categorias A5 e A6 de LOC (Tabela 15) e como houve um aumento do peso definido pelo utilizador (a categoria A5 de LOC passou a ter peso 1 ao invés do peso 0.8 atribuído anteriormente) a exactidão para estas classes aumentou. Mais uma vez a classe de vegetação herbácea (35), foi a que teve um decréscimo mais acentuado na exactidão (menos 15% que  $P_{LOC}$ ). Esta classe apresenta um elevado número de observações na classe A2 de LOC que não é



contemplada nesta análise definida pelo utilizador (peso 0 nesta categoria de LOC). Este aspecto evidencia mais uma vez, que o facto de esta classe estar presente na grande maioria das vezes em zonas em que a paisagem é heterogénea põe em evidência a incerteza que existe na sua interpretação aquando da elaboração de bases de dados de referência, advindo deste facto o seu reduzido valor de exactidão.

## 4. CONCLUSÃO

As medidas de exactidão  $P_{CONF}$  e  $P_{LOC}$ , pretendem ser medidas de exactidão temática que reflectem a influência que a incerteza na atribuição das classes de ocupação do solo de referência e a incerteza posicional das unidades amostrais de referência, podem ter na avaliação da exactidão temática de mapas de ocupação do solo. Neste estudo, a incerteza é incorporada na elaboração de bases de dados de referência de três maneiras: (1) através da atribuição de duas classes de ocupação do solo de referência; (2) através da atribuição de um índice de confiança na interpretação das classes de ocupação do solo de referência e; (3) através de um índice que traduz a possível influência que a incerteza posicional pode ter sobre a informação temática recolhida em cada uma das unidades amostrais de referência. Estes três tipos de incerteza são integrados no protocolo de avaliação da exactidão temática através das variáveis CONF que representa uma escala de confiança na classificação do mapa baseada na confiança da atribuição das classes de ocupação do solo de referência, e LOC que representa uma escala de confiança na classificação baseada na incerteza posicional.

Posteriormente é aplicado um esquema de pesos que reflecte as contribuições parciais de cada uma das categorias CONF e LOC para a exactidão temática do mapa em estudo. De facto, apresentar valores de exactidão temática que incorporem a incerteza na elaboração de bases de dados de referência, permite aos utilizadores de mapas de ocupação do solo, perceber os impactes que a incerteza tem nos índices de exactidão temática tradicionalmente utilizados (i.e. MAX e RIGHT). Para a implementação da metodologia de avaliação da exactidão temática proposta, é necessário recolher duas classes de ocupação do solo de referência, um índice de confiança de interpretação (ICR), e um índice de confiança de localização (LCR). Em outros estudos de avaliação da exactidão temática de mapas de ocupação do solo, esta informação foi recolhida sem um grande aumento de esforço no protocolo de avaliação da exactidão temática (Stehman *et al.* 2003 e Wickham *et al.* 2004). Tendo em conta que a recolha da informação de referência é um processo bastante moroso, a recolha desta informação adicional pode ser assim obtida sem grandes prejuízos de

tempo e possibilita a realização de possíveis análises que permitam uma melhor compreensão da exactidão temática de mapas de ocupação do solo.

Esta abordagem permite assim uma forma de os utilizadores finais de mapas de ocupação do solo verificarem o impacte que a incerteza na elaboração de bases de dados de referência tem nas medidas de exactidão, sendo este impacte reportado sob a forma da Tabela 14 (reflecte a incerteza na atribuição das classes de ocupação do solo de referência através de  $P_{CONF}$ ) e da Tabela 16 (incerteza posicional através de  $P_{LOC}$ ). O formato de ambas estas tabelas permite ao utilizador verificar que categorias de CONF ou LOC estão a contribuir para a exactidão final do mapa que pretendem utilizar. Para além disso, esta abordagem é bastante flexível e intuitiva, pois facilmente um utilizador pode alterar o esquema de pesos consoante as suas necessidades e recalcular os índices de exactidão temática propostos (i.e.  $P_{CONF}$  e  $P_{LOC}$ ).

No entanto este estudo levanta outras questões, nomeadamente como integrar a incerteza posicional e a incerteza na atribuição das classes de referência (Sarmiento *et al.* 2009a; Sarmiento *et al.* 2009b) e traduzir estes dois tipos de incerteza num único índice de exactidão temática. De facto a grande dificuldade prende-se com o esquema de pesos que deverá ser utilizado na elaboração deste índice, uma vez que a subjectividade torna-se bastante mais elevada, pois pretende-se englobar dois índices de incerteza (i.e. ICR e LCR). Para além disso é necessário também desenvolver uma metodologia que permita o cálculo dos intervalos de confiança das duas medidas de exactidão propostas neste estudo.

Outro aspecto que importa realçar é o desenvolvimento de métodos que retirem a subjectividade inerente ao esquema de pesos utilizado neste estudo, pois julgamos que só assim será possível desenvolver um método standardizado de avaliação da exactidão temática que incorpore a incerteza na elaboração de bases de dados de referência e que permita comparar a exactidão de mapas de ocupação do solo provenientes de várias fontes da mesma forma. Em estudos futuros e tendo em vista colmatar o aspecto anteriormente referido, pretende-se avaliar a magnitude do erro através da proximidade temática (i.e. um erro entre floresta e vegetação arbustiva é menos grave do que entre floresta e área agrícola) entre classes de ocupação do solo,

como proposto por Mayaux *et al.* (2006), através de uma matriz de pesos baseada no *Land Cover Classification System* (LCCS) (Di Gregorio e Jansen 2001). De facto este sistema de definição de nomenclaturas de ocupação do solo é amplamente aceite pela comunidade científica, como foi também o sistema utilizado para definir a nomenclatura utilizada neste estudo (i.e. nomenclatura LANDEO).

## REFERÊNCIAS

ARINO, O., LEROY, M., RANERA, F., GROSS, D., BICHERON, P., NINO, F., BROCKMAN, C., DEFOURNY, P., VANCUTSEM, C., ACHARD, F., DURIEUX, L., BOURG, L., LATHAM, J., DI GREGORIO, A., WITT, R., HEROLD, M., SAMBALE, J., PLUMMER, S., WEBER, J. L., GORYL, P. and HOUGHTON, N., 2007, GLOBCOVER - A Global Land Cover Service with MERIS. In *Proceedings of the ENVISAT Symposium*, 23-27 April 2007, Montreaux, Switzerland, unpaginated CD-ROM.

ARONOFF, S., 1982, Classification accuracy: A user approach. *Photogrammetric Engineering and Remote Sensing*, **48**, 1299-1307.

BROVKIN, V., SITCH, S., WERNER, V. B., CLAUSSEN, M., BAUER, E. and CRAMER, W., 2004, Role of land cover changes for atmospheric CO<sub>2</sub> increase and climate change during the last 150 years. *Global Change Biology*, **10**, 1253-1266.

CAETANO, M., MATA, F. and FREIRE, S., 2006, Accuracy assessment of the Portuguese CORINE Land Cover Map. In *Global Developments in Environmental Earth Observation from Space*, A. Marçal (Ed.), pp. 459-467 (Rotterdam: Millpress).

CARRÃO, H., ARAÚJO, A., CERDEIRA, C., SARMENTO, P., CAPÃO, L. and CAETANO, M., 2007, A reference sample database for the accuracy assessment of medium spatial resolution land cover products in Portugal. In *Proceedings of the IEEE International Conference on Geoscience and Remote Sensing Symposium*, 23-27 July 2007, Barcelona, Spain, unpaginated CD-ROM.

CARRÃO, H., GONÇALVES, P. and CAETANO, M., 2009, Multitemporal MERIS images for land cover mapping at national scale: the case study of Portugal. *International Journal of Remote Sensing* (in press).

CHAPIN, F. S. III, ZAVALA, E. S., EVINER, V. T., NAYLOR, R. L., VITOUSEK, P. M., REYNOLDS, H. L., HOOPER, D. U., LAVOREL, S., SALA, O. E., HOBBIE, S. E., MACK, M. C. and DIAZ, S., 2000, Consequences of changing biodiversity. *Nature*, **405**, 234-242.

CONGALTON, R. G. and GREEN, K. (Ed.), 1999, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, pp. 137 (Boca Raton: Lewis Publishers).

DI GREGORIO, A. and JANSEN, L. J. M., 2001, *Land Cover Classification System (LCCS): Classification Concepts and User Manual for Software Version 1.0*, pp. 194 (Rome: United Nations Food and Agricultural Organization).

EDWARDS, T. C., Jr., MOISEN, G. G. and CUTLER, D. R., 1998, Assessing map accuracy in an ecoregion-scale cover-map. *Remote Sensing of Environment*, **63**, 73-83.

FOODY, G. M., 2002, Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, **80**, 185-201.

GÓMEZ, D., BIGING, G. and MONTERO, J., 2008, Accuracy statistics for judging soft classification. *International Journal of Remote Sensing*, **29**, 693-709.

GOODCHILD, M., 2003, Geographic information science and systems for environmental management. *Annual Review of Environment and Resources*, **28**, 493-519.

GOPAL, S. and WOODCOCK, C., 1994, Theory and methods for accuracy assessment of thematic maps using fuzzy sets. *Photogrammetric Engineering and Remote Sensing*, **60**, 181-188.

HAGEN, A., 2003, Fuzzy set approach to assessing similarity between of categorical maps. *International Journal of Geographical Information Science*, **17**, 235-249.

LABA, M., GREGORY, S. K., BRADEN, J., OGURCAK, D., HILL, E., FEGRAUS, E., FIORE, J. and DEGLORIA, S. D., 2002, Conventional and fuzzy accuracy assessment of the New York Gap Analysis Project land cover maps. *Remote Sensing of Environment*, **81**, 443-455.

LANTER, D.P. and VEREGIN, H., 1992, A research paradigm for propagating error in layer-based GIS. *Photogrammetric Engineering and Remote Sensing*, **58**, 825-833.

LESSCHEN, J. P., VERBURG, P. H. and STAAL, S. J., 2005, *Statistical methods for analyzing the spatial dimension of changes in land use and farming systems*, pp. 80 (Nairobi and Wageningen: ILRI and University of Wageningen).

MAYAUX, P., EVA, H., GALLEG0, J., STRAHLER, A. H., HEROLD, M., AGRAWAL, S., NAUMOV, S., DE MIRANDA, E. E., DI BELLA, C. M., ORDOYNE, C., KOPIN, Y. and ROY, P. S., 2006, Validation of the Global Land Cover 2000 map. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 1728-1739.

MUCHONEY, D. M., STRAHLER, A. H., HODGES, J. and LOCASTRO, J., 1999, The IGBP DISCover confidence sites and the system for terrestrial ecosystem parametrization: Tools for validating global land-cover data. *Photogrammetric Engineering and Remote Sensing*, **65**, 1061-1067.

MULLER, S. V., WALKER, D. A., NELSON, F. E., AUERBACH, N. A., BOCKHEIM, J. G., GUYER, S. and SHERBA, D., 1998, Accuracy assessment of a land-cover map of the Kuparuk River Basin, Alaska: Considerations for remote regions. *Photogrammetric Engineering and Remote Sensing*, **64**, 619-628.

SARMENTO, P., 2007, Avaliação da exactidão temática de cartografias de ocupação do solo: Abordagem aplicada a mapas derivados de imagens de satélite de média resolução espacial. Tese de licenciatura, Universidade de Évora, Portugal.

SARMENTO, P., CARRÃO, H. e CAETANO, M., 2008a, Avaliação da exactidão temática de cartografias de ocupação do solo através de funções *fuzzy*: primeira abordagem. In *Actas do X Encontro de Utilizadores de Informação Geográfica*, 14-16 Maio 2008, Oeiras, Portugal.

SARMENTO, P., CARRÃO, H. e CAETANO, M., 2008b, A fuzzy synthetic evaluation approach for land cover cartography accuracy assessment. In *Spatial Uncertainty*, J. Zhang and M. Goodchild (Ed.), pp. 348-355 (Liverpool: World Academic Union).

SARMENTO, P., CARRÃO, H., CAETANO, M. and STEHMAN, S. V., 2009a,

Incorporating reference classification uncertainty into the analysis of land cover accuracy. *International Journal of Remote Sensing* (in press).

SARMENTO, P., FONTE, C. e CAETANO, M., 2009b, Integração da incerteza posicional de bases de dados de referência na análise da exactidão temática de mapas de ocupação do solo. In *Actas da VI Conferência Nacional de Cartografia e Geodesia (CNCG 2009)*, 7-8 Maio 2009, Caldas da Rainha, Portugal.

STEHMAN, S. V., 1997, Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, **62**, 77-89.

STEHMAN, S. V. and CZAPLEWSKI, R. L., 1998, Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of Environment*, **64**, 331-344.

STEHMAN, S. V., WICKHAM, J. D., SMITH, J. H. and YANG, L., 2003, Thematic accuracy of the 1992 National Land-Cover Data for the eastern United States: Statistical methodology and regional results. *Remote Sensing of Environment*, **86**, 500-516.

STORY, M. and CONGALTON, R. G., 1986, Accuracy assessment: A user's perspective. *Photogrammetric Engineering & Remote Sensing*, **52**, 397-399.

STRAHLER, A. H., BOSCHETTI, L., FOODY, G. M., FRIEDL, M. A., HANSEN, M. C., HEROLD, M., MAYAUX, P., MORISSETTE, J. T., STEHMAN, S. V. and WOODCOCK C. E., 2006, *Global Land Cover Validation: Recommendations for Evaluation and Accuracy Assessment of Global Land Cover Maps*, Report of Committee of Earth Observation Satellites (CEOS) - Working Group on Calibration and Validation (WGCV).

VITOUSEK, P. M., 1994, Beyond global warming: ecology and global change. *Ecology*, **75**, 1861-1876.

WICKHAM, J. D, STEHMAN, S. V., SMITH, J. H. and YANG, L., 2004, Thematic accuracy of the 1992 National Land-Cover Data for the western United States. *Remote Sensing of Environment*, **91**, 452-468.



WOODCOCK, C. and GOPAL, S., 2000, Fuzzy set theory and thematic maps: accuracy assessment and area estimation. *International Journal of Geographical Information Science*, **14**, 153-172.

WOODCOCK, C., GOPAL, S. and ALBERT, W., 1996, Evaluation of the potential for providing secondary labels in vegetation maps. *Photogrammetric Engineering and Remote Sensing*, **62**, 393-399.

YANG, L., STEHMAN, S. V., SMITH, J. H. and WICKHAM, J. D., 2001, Thematic accuracy of MRLC land cover for the eastern United States. *Remote Sensing of Environment*, **76**, 418-422.

YANG, L., STEHMAN, S. V., WICKHAM, J. D., JONATHAN, S. and VAN DRIEL, N. J., 2000, Thematic validation of land cover data of the eastern United States using aerial photography: Feasibility and challenges. In *4<sup>th</sup> International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, 12-14 July 2000, Amsterdam (Delft University Press), pp. 747-754.

ZHU, Z., YANG, L., STEHMAN, S. V. and CZAPLEWSKI, R. L., 2000, Accuracy assessment from the US Geological Survey regional land cover mapping program: New York and New Jersey region. *Photogrammetric Engineering and Remote Sensing*, **66**, 1425-1435.